ROGER PENROSE

Tehnos

INCERTITUDINILE RAȚIUNII

Umbrele minții

Editura Tehnică

Coordonatorul colecției: dr. Roman Chirilă

τ

Există, fără îndoială, un soi de fericire căpătată din înțelegerea celor privite ori gândite, un soi de împlinire fecundă derivată din efortul de a descifra misterele realității.

Colecția τ (Taifas) își propune să găzduiască acele cărți care să-i prilejuiască cititorului întâlnirea sa cu Universul, cu lumea înconjurătoare, prin mijlocirea cuvântului tăifăsuit; cărți prin care natura lucrurilor este adusă în dezbatere critică, obiectivă de către mari spirite creatoare și modelatoare ale științei fără formule, fără matematică și fără ca reprezentările-imagini despre Lume să fie alterate; cărți prin care tăifăsuirea să devină o ipostază a cunoașterii științifice și o re-cunoaștere a ordinii prealabile a Universului.

Cuvântul tăifăsuit face posibil orice excurs mental, imaginar, confrerizând diversități, proximizând îndepărtări și revelând înțelesuri.

Cititorul instruit pentru o astfel de lectură, trudnică și răbdătoare, va deveni curând beneficiarul unui spor cognitiv indiscutabil, capabil de a îi potența nostalgia și sentimentul infinitului, a vastității lumii înconjurătoare și a plenitudinii ei coerente, a mirării ca sentiment al înălțării sale interioare !...

ROGER PENROSE

NCERTITUDINILE RAȚIUNII

(UMBRELE MINȚII)

ÎN CĂUTAREA UNEI TEORII ȘTIINȚIFICE A CONȘTIINȚEI

Traducere din limba engleză
[illegible]

EDI
București, 1999

Copyright © 1998, Editura TEHNICA
Toate drepturile asupra acestei ediţii sunt rezervate editurii

Adresa: S.C. Editura TEHNICĂ S.A.
Piaţa Presei Libere, nr.1
33 Bucureşti, România
Cod 71341

© Roger Penrose 1994
„This translation of *Shadows of the Mind* originally published in English in 1994 is published by arrangement with Oxford University Press".
Această traducere a lucrării *Shadows of the Mind* a fost efectuată după ediţia originală apărută în limba engleză în 1994 şi este publicată prin înţelegerea cu Oxford University Press.

Editată cu spijinul Ministerului Cercetării şi Tehnologiei

Coperta colecţiei: **Andrei Mănescu**

Redactor: **Denisa Popa**
Tehnoredactor: **Diana Jilavu**

Bun de tipar: 20.12.1998; Coli tipo: 33,5
C.Z.U.: 159.922

ISBN 973-31-1290-9

Tipărit la Tipografia SEMNE

CUPRINS

Prefață 11

Nota traducătorului 17

Prolog 19

Partea întâi:
De ce avem nevoie de o nouă fizică pentru a înțelege mintea 23

1. **Conștiința și calculul** 25
   - 1.1. Conștiință și știință 25
   - 1.2. Pot roboții salva această lume tulburată? 27
   - 1.3. $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$ –ul calculului și gândirii conștiente 31
   - 1.4. Fizicalism și mentalism 37
   - 1.5. Calcul: proceduri top-down și bottom-up 38
   - 1.6. Poziția $\mathcal{C}$ contravine tezei Church-Turing? 41
   - 1.7. Haos 43
   - 1.8. Calcul analogic 46
   - 1.9. Ce fel de acțiune ar putea fi non-computațională? 48
   - 1.10. Ce se poate spune despre viitor? 56
   - 1.11. Pot calculatoarele să aibă drepturi și responsabilități? 58
   - 1.12. 'Conștiență', 'înțelegere', 'conștiință', 'inteligență' 60
   - 1.13. Argumentul lui John Searle 65
   - 1.14. Unele dificultăți ale modelului computațional 66
   - 1.15. Limitările actuale din IA oferă sprijin poziției $\mathcal{C}$? 69

1.16. Argumentul oferit de teorema lui Gödel 74
1.17. Platonism sau misticism ? 76
1.18. Cât de relevantă este înţelegerea matematică? 77
1.19. Ce legătură are teorema lui Gödel cu comportamentul de zi cu zi ? 79
1.20. Vizualizarea mentală şi realitatea virtuală 84
1.21. Este imaginaţia matematică nealgoritmică ? 86
Note şi referinţe bibliografie 89

2. Exemplul lui Gödel **93**
2.1. Teorema lui Gödel şi maşinile Turing 93
2.2. Proceduri de calcul 95
2.3. Calcule care nu se termină 97
2.4. Cum să decidem dacă un anumit calcul se termină sau nu ? 98
2.5. Familii de calcule; concluzia Gödel-Turing ($\mathcal{G}$) 103
2.6. Obiecţiile tehnice posibile relative la $\mathcal{G}$ 108
2.7. Câteva consideraţii matematice mai profunde 122
2.8. Condiţia de ω-consistenţă 125
2.9. Sistemele formale şi demonstraţia algoritmică 128
2.10. Alte obiecţii tehnice posibile relativ la $\mathcal{G}$ 131
Note şi referinţe bibliografice 156

Anexa A O maşină Turing de gödelizare 158

3. Despre non-computabilitate în gândirea matematică **168**
3.1. Ce au gândit Gödel şi Turing ? 168
3.2. Poate un algoritm incorect să simuleze înţelegerea matematică într-un mod care poate fi cunoscut? 172
3.3. Poate simula înţelegerea matematică un algoritm cognoscibil într-un mod care nu poate fi cunoscut? 174
3.4. Utilizează oare, fără discernământ, matematicienii un algoritm incorect ? 182
3.5. Poate un algoritm să fie incognoscibil? 187
3.6. Selecţie naturală sau acţiunea lui Dumnezeu? 190
3.7. Un algoritm sau mai mulţi algoritmi? 191
3.8. Selecţia naturală a matematicienilor 193
3.9. Algoritmi de învăţare 197

3.10. Poate mediul să fie un factor extern nealgoritmic? 199
3.11. Cum poate un robot să învețe? 201
3.12. Poate un robot să aibă 'convingeri matematice ferme'? 204
3.13. Mecanisme care stau la baza matematicii robotului 208
3.14. Contradicția fundamentală 211
3.15. Mijloace prin care contradicția poate fi îndepărtată 213
3.16. Trebuie robotul să creadă în M ? 214
3.17. Erorile și 'înțelegerea' robotului 218
3.18. Cum poate fi încorporat elementul aleator în activitatea robotului? 219
3.19. Suprimarea aserțiunilor-★ eronate 221
3.20. Este nevoie doar de un număr finit de aserțiuni-★$_{\mathcal{M}}$. 224
3.21. Sunt suficiente măsurile de prevedere luate? 228
3.22. Poate haosul să salveze modelul computațional al minții? 230
3.23. *Reductio ad absurdum* un dialog fantastic 231
3.24. Raționamentul pe care l-am făcut este cumva paradoxal? 245
3.25. Complicația în demonstrațiile matematice 249
3.26. Întreruperea prin calcul a ciclurilor infinite 252
3.27. Matematică computațională *top-down* sau *bottom-up*? 256
3.28. Concluzii 259
Note și referințe bibliografice 267

Partea a doua:
Ne trebuie o nouă fizică pentru a înțelege mintea 271

4. Are mintea un loc în fizica clasică? 273
4.1. Mintea și legile fizicii 273
4.2. Calculabilitate și haos în fizica zilelor noastre 275
4.3. Conștiința: o nouă fizică sau un 'fenomen emergent'? 277
4.4. *Înclinarea* Einstein 278
4.5. Calcul și fizică 290
Note și referințe bibliografice 299

5. Structura lumii cuantice **300**
5.1. Teoria cuantică – enigmă şi paradox 300
5.2. Problema Elitzur-Vaidman de testare a bombelor 302
5.3. Dodecaedrele magice 304
5.4. Statutul experimental al misterelor Z de tip EPR 310
5.5. Piatra de temelie a teoriei cuantice – o istorie a extraordinarului 314
5.6. Regulile de bază ale teoriei cuantice 322
5.7. Evoluţia unitară (U) 325
5.8. Reducerea vectorului de stare **R** 329
5.9. Soluţia problemei lui Elitzur şi Vaidman 334
5.10. Teoria cuantică a spinului. Sfera Riemann 337
5.11. Poziţia şi impulsul unei particule 346
5.12. Spaţiul Hilbert 347
5.13. Descrierea lui **R** cu ajutorul spaţiului Hilbert 351
5.14. Măsurători comutative 355
5.15. "şi"-ul mecanicii cuantice 356
5.16. Ortogonalitatea stărilor produs 359
5.17. Nonlocalizarea cuantică 360
5.18. Explicaţia dodecaedrelor magice 366
Anexa B: Imposibilitatea de a colora dodecaedrul 372
Anexa C: Ortogonalitate între stările generale de spin 373
Note şi referinţe bibliografice 377

6. Teoria cuantică şi realitatea **379**
6.1. Este **R** un proces real? 379
6.2. Puncte de vedere de tip *many worlds* 383
6.3. Neluându-l pe $|\psi>$ în serios 386
6.4. Matricea densitate 390
6.5. Matrice densitate pentru perechi EPR 395
6.6. O explicaţie FAPP a lui **R**? 398
6.7. Explică FAPP regula modulului pătrat? 403
6.8. Este conştiinţa cea care reduce vectorul de stare? 404
6.9. Luând $|\psi>$ în serios 406
6.10. Reducţie a vectorului de stare indusă gravitaţional? 410
6.11. Unităţi absolute 414
6.12. Noul criteriu 415
Note şi referinţe bibliografice 423

7. Teoria cuantică şi creierul **425**
7.1. Acţiune cuantică la scară mare în funcţionarea creierului? 425
7.2. Neuroni, sinapse şi calculatoare 430
7.3. Calculabilitate cuantică 434
7.4. Citoschelete şi microtubule 436
7.5. Coerenţă cuantică în microtubule? 447
7.6. Microtubulele şi conştiinţa 450
7.7. Un model pentru minte? 452
7.8. Non-computabilitate în gravitaţia cuantică (1) 459
7.9. Maşini oracol şi legi fizice 462
7.10. Non-computabilitate în gravitaţia cuantică (2) 465
7.11. Timpul şi percepţiile conştiente 467
7.12. EPR şi timpul: necesitatea unei noi imagini a lumii 473
Note şi referinţe bibliografice 476

8. Implicaţii? **478**
8.1. 'Aparate' artificiale inteligente 478
8.2. Lucruri pe care calculatoarele le fac bine sau prost 481
8.3. Estetică şi altele 485
8.4. Anumite pericole inerente în tehnologia calculatoarelor 487
8.5. Alegerile enigmatice 488
8.6. Fenomenul fizic al conştiinţei? 491
8.7. Trei lumi şi trei mistere 497
Note şi referinţe bibliografice 508

Epilog **509**

Bibliografie 510

Sir Roger Penrose scurtă notă biografică 531

Termeni propuşi în limba română 532

PREFAȚĂ

Cartea de față poate fi privită, într-un anumit sens, ca o continuare a precedentei mele cărți „The Emperor's New Mind" (la care mă voi referi cu prescurtarea ENM)[1]. Într-adevăr, voi continua aici tema inițiată în ENM, dar cartea poate fi citită complet independent. Motivația de a scrie din nou despre același subiect se datorează în parte nevoii de a răspunde detaliat unui mare număr de întrebări și observații critice pe care diferite persoane le-au făcut în legătură cu argumentele expuse în ENM. Acum voi explora idei noi care depășesc pe cele din ENM. Una dintre temele centrale din ENM a fost părerea mea că prin utilizarea propriei conștiințe suntem capabili să îndeplinim acțiuni care se află deasupra oricărui fel de activitate computațională[2] Totuși, în ENM această idee a fost prezentată ca o ipoteză și a rămas oarecum neprecizat care proceduri ar putea fi incluse în ceea ce numim 'activitate computațională'. Prezentul volum oferă ceea ce eu cred că este un studiu mult mai riguros care conduce la aceeași concluzie și care se aplică oricărui tip de proces computațional. Mai mult, prezint aici o descriere mult mai plauzibilă, decât putea fi dată în ENM, pentru funcționarea creierului conform căreia o acțiune fizică non-computațională ar putea susține comportarea noastră comandată conștient.

Studiul meu se desfășoară pe două direcții distincte. Prima este în mod esențial negativă și se referă la atitudinea mea hotărâtă împotriva punctului de vedere comun prin care activitatea mentală conștientă, în toate manifestările sale variate, ar putea fi, în principiu, înțeleasă cu ajutorul unor modele computaționale. Cealaltă direcție a demersului meu este pozitivă în sensul că reprezintă o căutare,

---

[1] Traducerea în limba română a apărut cu titlul „Mintea noastră...cea de toate zilele", în traducerea lui Mircea Rusu și a Corneliei Rusu, la Editura Tehnică, în 1996.(n.t.)

[2] Prin activitate computațională desemnăm activitatea care poate fi realizată de un calculator.(n.t.)

în interiorul constrângerilor impuse de o tratare științifică, a mijloacelor prin care un creier descris științific ar putea să folosească principiile subtile ale fizicii, unele încă necunoscute, pentru a executa acțiunile non-computaționale.

În concordanță cu această dihotomie, argumentele din această carte sunt prezentate în două părți separate. Partea întâi conține o discuție detaliată și completă, susținând cu putere teza mea că 'înțelegerea' omenească, ca manifestare particulară a conștiinței face ceva ce nu poate fi făcut de nici un sistem de calcul. Precizez că termenul 'sistem de calcul' include atât sistemele *top-down* care acționează conform unor pocese algoritmice bine precizate și înțelese cât și sistemele *bottom-up* care sunt programate în așa fel încât să permită învățarea din experiență. Partea întâia a cartii are în centru faimoasa teoremă a lui Gödel și conține o examinare detaliată a implicațiilor relevante ale teoremei lui Gödel. Această examinare extinde mult argumentele inițiale furnizate de Gödel însuși, Nagel și Newman și de către Lucas și răspunde amănunțit diferitelor obiecții critice de care am luat cunoștință. În legătură cu aceasta sunt aduse argumente temeinice împotriva ideii că sistemele *bottom-up* (ca și cele *top down*) ar fi capabile să atingă vreodată inteligența pură. Concluzia este că gândirea conștientă trebuie să conțină ingrediente care nu pot fi măcar simulate adecvat prin calcul pur și simplu; cu atât mai puțin poate un sistem de calcul, el singur, să manifeste vreun sentiment sau intenție conștientă. Înseamnă că mintea nu poate fi adecvat descrisă, în nici un fel, prin metode computaționale.

În partea a doua, argumentația se axează pe fizică și biologie. Linia raționamentului, deși conține porțiuni care sunt în mod evident mai mult propuneri decât o prezentare riguroasă ca cea din partea întâi, reprezintă o încercare originală de a înțelege modul în care o acțiune non-computațională ar putea să apară în cadrul legilor fizicii comprehensibile științific. Sunt introduse din nou principiile de bază ale mecanicii cuantice — ele nefiind necesare cititorului care are deja cunoștințe de mecanică cuantică. Dar, în același timp, enigmele, paradoxurile și misterele acestui subiect sunt adâncite, utilizând un număr de exemple noi care ilustrează grafic rolul important al non-localizării și contrafactualității și implicațiile adânci pe care ciudatele fenomene cuantice le ridică. Voi aduce argumente serioase pentru nevoia de schimbare fundamentală în imaginea lumii din mecanica cuantică, precizând cu claritate nivelul la care această schimbare este necesară. (Aceste idei sunt strâns legate de lucrările recente ale lui Ghirardi, Diósi și alții.) Există diferențe importante între ideile prezentate aici și cele susținute în ENM.

Sunt de părere că o non-computabilitate fizică este necesară pentru a explica non-computabilitatea din activitatea conștientă. Ca urmare, voi susține ca nivelul la care non-computabilitatea fizică este semnificativă, trebuie să aibă importanță în

funcționarea creierului. Aici este punctul în care propunerile mele actuale diferă substanțial de cele prezentate în ENM. Sunt de părere că în vreme ce semnalele neuronale se pot comporta ca evenimente clasice determinate, legăturile sinaptice dintre neuroni sunt comandate la un nivel mai adânc unde este de așteptat să existe o activitate fizică importantă la granița cuantic-clasic. Presupunerile pe care le fac cer să existe o comportare cuantic-coerentă, la scară mare, (în concordanță cu propunerile făcute de Fröhlich) intervenind în interiorul microtubulelor din citoscheletonul neuronilor. Impresia este că această activitate cuantică trebuie să fie legată non-computațional de activitatea de tip computațional pe care a prezentat-o Hameroff și colegii săi ca având loc de-a lungul microtubulelor.

Argumentele pe care le prezint trimit spre câteva puncte în care imaginea științifică actuală se dovedește profund neputincioasă să ofere o înțelegere științifică a activității minții umane. Dar fără îndoială, acest lucru nu înseamnă că fenomenul conștiinței ar trebui să rămână în afara cercetării științifice. Susțin cu tărie, așa cum am făcut-o și în ENM că ar trebui să existe o metodă științifică de înțelegere a fenomenelor mentale și că aceasta ar trebui să înceapă cu o înțelegere mai adâncă a naturii realității fizice însăși. Cred că este important ca orice cititor preocupat de subiect, dorind să pătrundă modul cum un fenomen atât de enigmatic, cum este mintea, poate fi înțeles în contextul unei lumi fizice materiale, va trebui să accepte regulile stranii care trebuie să guverneze, de fapt, acest 'material' al lumii noastre fizice.

Înțelegerea este, până la urmă, scopul științei - iar știința este mult mai mult decât un calcul mecanic, lipsit de înțelegere.

Oxford Roger Penrose
Aprilie 1994

Mulțumiri

Pentru sprijinul acordat în scrierea acestei cărți sunt multe persoane cărora le sunt foarte îndatorat - prea multe ca să le pot mulțumi individual, chiar dacă mi-aș aminti numele tuturor. Totuși, sunt în special recunoscător lui Guido Bacciagaluppi și Jeremy Butterfield pentru observațiile critice în anumite părți ale unei prime versiuni unde ei au descoperit o eroare ascunsă în raționamentul de atunci, în ceea ce acum a devenit capitolul 3 al cărții. Sunt recunoscător de asemenea lui Abhay Ashtekar, Bryan Birch, Geoff Brooker, David Chalmers, Francis Crick, David Deutsch, Susan Greenfield, Andrew Hodges, Dipankar Home, Ezio Insinna, Dan Isaacson, Roger James, Richard Jozsa, John Lucas, Bill McColl, Claus Moser, Graeme Michison, Ted Newman, Oliver Penrose, Jonathan Penrose, Stanley Rosen, Ray Sachs, Graeme Segal, Aaron Sloman, Lee, Smolin, Ray Streater, Valerie Willoughby, Anton Zeilinger, și în special lui Artur Ekert pentru informații variate si ajutor. Au existat nenumărați corespondenți și persoane care mi-au oferit verbal comentariile lor la cartea precedentă *The Emperor's New Mind*. Le mulțumesc aici chiar dacă cei mai mulți dintre ei mai așteaptă încă răspunsurile mele la scrisorile lor! Dacă n-aș fi beneficiat de punctele lor de vedere privind precedenta mea carte, este prea puțin probabil că m-aș fi angajat la munca descurajantă de a mai scrie una.

Sunt recunoscător organizatorilor Messenger Lectures de la Universitatea Cornell (unde am prezentat conferințele cu titlul secțiunii finale a acestei cărți), organizatorilor Gifford Lectures de la Universitatea din St. Andrews, Forder Lectures din Noua Zeelandă, Gregynogg Lectures de la Universitatea din Aberystwyth, precum și a unei serii de lecții distinse de la Five Colleges, Amherst, Massachusetts pe lângă nenumărate conferințe în diferite colțuri ale lumii.

Acestea au fost ocazii care mi-au oferit posibilitatea de a-mi împrospăta ideile și a obține reacții prețioase din partea audienței. Mulțumesc institutului Isaac Newton din Cambridge, precum și Universității Syracuse și Universității Penn State pentru onoarea de a mă premia, respectiv de a-mi oferi titlul de *Visiting Distinguished Professorship in Mathematics and Physics*, și *Francis R. Pentz and Helen M. Pentz Distinguished Professorship in Physics and Mathematics*. Mulțumesc de asemenea Fundației Naționale pentru știință pentru sprijinul acordat prin contractele PHY 86-12424 și PHY 43-96246.

În sfârșit mai sunt trei persoane care merită în mod special să fie amintite. Mi-au fost de un enorm ajutor asistența și sprijinul dezinteresat al lui Angus MacIntyre pentru verificarea argumentelor de logică matematică din capitolele 2 și 3 oferindu-mi multe referințe care mi-au fost de un real folos. Îi adresez cele mai calde mulțumiri. Stuart Hameroff m-a învățat despre structura de citoscheleton și microtubulele sale, despre care, cu doi ani în urmă, nici nu știam că există. Îi mulțumesc mult pentru inestimabilele informații și pentru ajutorul dat la verificarea celei mai mari părți din capitolul 7. Îi voi rămâne veșnic îndatorat pentru felul în care mi-a deschis ochii spre minunile unei lumi noi. El, ca și ceilalți cărora le mulțumesc, nu este în nici un fel responsabil pentru erorile care, nu mă îndoiesc, au rămas în această carte. Mai mult și mai mult, dragei mele Vanessa, i se datorează mulțumiri speciale din mai multe motive: pentru că mi-a explicat de ce trebuiau rescrise anumite părți din carte, pentru răbdarea ei și pentru înțelegere, în special atunci când am subestimat timpul necesar pentru scris. Și, ei bine, îi mulțumesc pentru că, fără să știe a oferit *imaginației* mele modelul Jessicăi și povestea acesteia. Regret doar că nu am cunoscut-o la vârsta Jessicăi.

Mulțumiri pentru figuri

Editorii ediției originale mulțumesc următorilor pentru permisiunea de a reproduce material ilustrativ:

Fig. 1.1.	A. Nieman / Science Photo Library
Fig. 4.12.	din J.C. Mather *et al.* (1990), *Astrophys. J.*, **354**, L37
Fig. 5.7.	din A. Aspect and P. Grangier (1986), din *Quantum concepts in space and time* (ed. R. Penrose and C.J. Isham), p. 1-27, Oxford University Press
Fig. 5.8.	din Ashmolean Museum, Oxford
Fig. 7.2.	din R. Wichterman (1986), *The biology of paramecium*, 2nd edn. Plenum Press, New York
Fig. 7.6.	Eric Grave /Science Photo Library
Fig. 7.7.	din H. Weyl (1943), *Symmetry*, 1952 Princeton University Press
Fig. 7.10.	din N. Hirokawa (1991), din *The neuronal cytoskeleton* (ed. R.D. Burgoyne), p. 5-74, Wiley Liss, New York

Către cititor

Anumite părţi din această carte diferă mult de altele în privinţa gradului de dificultate. Părţile cele mai complicate din punct de vedere tehnic cred că sunt Anexele A şi C. Nu va fi însă nici o pierdere pentru cititor dacă le va ignora, pur şi simplu. Acelaşi lucru se poate afirma despre părţile mai tehnice din capitolele 2 şi 3. Acestea se adresează în principal cititorilor care trebuiesc convinşi de tăria cu care lupt împotriva oricărui model pur computaţional al înţelegerii umane. Pe de altă parte, cititorii deja convinşi (sau grăbiţi) pot prefera un drum mai puţin neplăcut către esenţa acestei cărţi. Acest drum se obţine citind, pur şi simplu, dialogul fantastic din paragraful 3.23, de preferinţă precedat de capitolul 1 şi paragrafele 2.1-2.5 şi 3.1.

Matematica cea mai serioasă din această carte este în legătură cu prezentarea mecanicii cuantice. Se referă în special la spaţiile Hilbert descrise în paragrafele 5.12-5.18 şi, în special, în discuţiile din paragrafele 6.4-6.6 centrate pe matricea densitate - importantă pentru a aprecia de ce am nevoie de o teorie *îmbunătăţită* a mecanicii cuantice! Sfatul meu pentru cititorul nematematician (sau chiar pentru matematician) este ca, atunci când întâlneşte expresii matematice înfricoşătoare să treacă peste ele, când constată că o examinare mai atentă nu aduce o înţelegere suplimentară. Este adevărat că subtilităţile mecanicii cuantice nu pot fi apreciate pe deplin fără a face cunoştinţă cu matematica ei elegantă şi misterioasă; totuşi, ceva din parfumul subiectului va răzbi fără îndoială chiar dacă matematica sa este complet ignorată.

În plus, trebuie să cer iertare cititorului şi dintr-un alt motiv. Sper ca el sau ea să mă ierte atunci când mă voi referi într-un mod despre care s-ar putea presupune că implică sexul. Totuşi, în discursul utilizat frecvent în carte, mă refer adeseori la o persoană abstractă, observator sau fizician. Este evident că nu apare nici o implicaţie în privinţa sexului individului, dar limba engleză nu permite folosirea pronumelui personal (persoana treia singular) la genul neutru, iar utilizarea repetată a expresiei „el sau ea"[3] este cu siguranţă obositoare.

Ca o observaţie finală, trebuie să precizez că referinţele la *The Emperor's New Mind* privesc ediţia[4] originală Oxford University Press, New York (hardback).

---

[3] Cum în limba română nu este uzual acest mod de exprimare, traducătorii nu au avut această problemă.(n.t)

[4] Versiunea prezentă, în limba română, face trimiteri la „Mintea noastră cea de toate zilele", Editura Tehnică, Bucureşti 1996 (titlu abreviat în text cu MN). (n.t.)

Nota traducătorului

Autor a mai multor cărţi de succes, Roger Penrose a ajuns un nume cunoscut cu mult dincolo de comunitatea ştiinţifică al cărei reprezentant de frunte este de vreo treizeci de ani. Cărţile lui sunt în acelaşi timp o demonstraţie şi o provocare. O demonstraţie a faptului că, un volum în care este vorba despre teorii cosmologice contemporane, matematici aplicate, inteligenţă artificială, neurobiologie şi alte câteva lucruri - toate într-un riguros limbaj ştiinţific - poate ajunge *best-seller*, poate stârni discuţii mult dincolo de domeniile citate, poate produce ecouri în mass media şi deveni subiect de controversă între filozofi. În acelaşi timp, Roger Perose adresează prin intermediul cărţilor sale o continuă provocare comunităţii ştiinţifice. Credeţi că ştiinţa a rezolvat deja, în mare, toate problemele esenţiale cu care omenirea se confruntă? Credeţi că până la o „mare teorie unificată" în fizica mai avem doar un pas ? Credeţi că neurobiologia, psihologia cognitivă sau chiar filozofia minţii sunt în stare să ne dea o teorie ştiinţifică a felului în care mintea noastră este alcătuită ? Sunteţi atunci dintre cei cărora provocarea le este adresată.

Pentru că Roger Penrose este convins că suntem încă destul de departe de toate acestea; că în felul nostru de a gândi şi de a ne construi teoriile trebuie să se producă o schimbare radicală înainte de a putea spera să răspundem la câteva dintre întrebările cele mai simple. Ce este mintea? Este ea mai mult dacât creierul şi dacă da, în ce sens? Cum funcţionează gândirea? Ce sunt raţiunea, percepţiile, sentimentele noastre, privite din punct de vedere *ştiinţific* ?

În volumul de faţă. Penrose reia problematica din „The Emperor's New Mind", („Mintea noastră... cea de toate zilele" apărut la Editura Tehnică în 1996 în traducerea lui Mircea şi a Corneliei Rusu), în dorinţa de a răspunde numeroşilor săi oponenţi cu o abordare mai riguroasă. Aici, argumentele autorului sunt concentrate, pe de o parte, într-o demonstraţie de principiu - bazată pe transpunerea teoremei lui Gödel, iar pe de altă parte, pe analiza elementelor care lipsesc din ştiinţa actuală din fizică şi biologie care să permită studiul ştiinţific al conştiinţei. Apar astfel noi probleme „fierbinţi" care nu pot fi circumscrise unui singur domeniu.

Din această situare, la graniţa dintre domeniile pe care autorul le transgresează mereu, apar şi problemele traducătorului. O foarte mare parte din termenii „tehnici" deja tipici pentru domenii de cercetare consacrate nu există în limba română sau, recent introduşi, nu au încă acordul deplin al specialiştilor. Să amintim numai „filosofia minţii" (*Philosophy of the mind*) care mai stârneşte încă rezervele multora, mai ales dacă vin dinspre biofizică sau neurobiologie. Situaţii similare am întâlnit aproape la fiecare pagină, termeni provenind din mecanica cuantică, cosmologie sau inteligenţa artificială. Să subliniem deci că soluţiile propuse în această traducere sunt provizorii - am încercat să le însoţim cu o notă de subsol care să conţină termenul în englezeşte şi eventualele soluţii alternative, iar câţiva termeni din această categorie i-am grupat într-un tabel la p. 532.

Pe de altă parte, au apărut şi alte dificultăţi. Aflat el însuşi în încurcatură, mereu la frontierele cunoaşterii actuale, Roger Penrose se vede nevoit să introducă termeni noi, să împrumute, schimbând contextul, termeni vechi, să utilizeze metafore şi aşa mai departe. Numărul de cuvinte între ghilimele este considerabil; se poate vedea în el o încercare a autorului de a spune lucrurile şi altfel, tot aşa cum poate sta mărturie pentru faptul că o mare schimbare în ştiinţa modernă este în pregătire.

Nu în ultimul rând, titlul însuşi este cel care ridică probleme: „Shadows of the Mind" (Umbrele Minţii), trimite la o imagine incitantă şi lămuritoare pe care o veţi găsi în prolog. Roger Penrose rescrie mitul platonician al peşterii cu aceeaşi dezinvoltură şi cu acelaşi umor de bună calitate pe care îl puteţi găsi pe tot parcursul cărţii.*

Căci, până la urmă, tot ce putem şti despre varianta modernă a întrebării „cine suntem ?" se reduce la o teorie (riguroasă !) care descrie umbrele lăsate pe peretele peşterii de un obiect despre care ştim prea puţin.

Şi încă o observaţie, care ar putea suna aşa: nu vă speriaţi de formule. La prima vedere, cartea lui Penrose pare o carte de popularizare a ştiinţei. La o a doua privire, poate părea o carte complicată, hiperformalizată ş.a.m.d. Aş spune că ambele judecăţi sunt eronate şi că o privire ceva mai atentă nu va întârzia să descopere deliciile unei abordări entuaziaste, imaginative şi pline de umor - însă imaginaţia, entuziasmul şi umorul unuia dintre cei mai mari savanţi ai secolului nostru.

Bucureşti, decembrie 1998 Dana Jalobeanu

---

* Titlul *Incertitudinile raţiunii* a fost ales din considerente filologice (N.E.)

PROLOG

Jessica se simţea întotdeauna puţin nervoasă când intra în partea asta a peşterii. „Tati, dacă piatra aia mare cade din locul în care stă? Nu-i aşa că asta ne-ar bloca ieşirea şi n-am mai putea ajunge niciodată, niciodată acasă?

„Aşa ar fi, dar n-o să cadă", răspunse tatăl ei puţin absent şi destul de brusc preocupat de modul în care plantele lui începeau să se obişnuiască cu întunericul în această parte, cea mai întunecată, a peşterii.

„Dar de unde ştii că n-o să cadă, tati?" insistă Jessica.

„Stânca aia e probabil acolo de multe mii de ani. N-o să cadă tocmai acum."

Jessica nu era satisfăcută cu atât. „Cu siguranţă că, dacă e să cadă odată, cu cât a stat mai mult acolo sus cu atât e mai mare pericolul să cadă acum, nu?"

Tatăl Jessicăi se opri din examinarea plantelor şi se uită la ea zâmbind uşor. „Nu e deloc aşa" Zâmbetul deveni mai vizibil, apoi mai interiorizat. „De fapt, cu cât stă acolo de mai multă vreme, cu atât e mai *puţin* probabil că va cădea cât timp suntem noi aici." Era evident că nu avea să urmeze nici un fel de explicaţie; tatăl se întoarse la plantele lui.

Jessica nu-l putea suferi când era într-o asemenea dispoziţie dar, de fapt, nu era chiar aşa, îl iubea foarte mult, mai mult decât pe oricine altcineva. Ar fi dorit, însă, să nu aibă dispoziţii de acest fel. Ştia că toate astea au ceva de-a face cu faptul că tatăl ei era un savant, dar tot nu înţelegea. Jessica spera să devină şi ea savant într-o zi, dar va avea grijă să nu aibă niciodată o astfel de dispoziţie.

Dar cel puţin se liniştise în privinţa pericolului ca bolovanul să se rostogolească şi să blocheze ieşirea. Era evident că pe tatăl ei nu-l neliniştea această posibilitate, iar încrederea lui îi alungă nervozitatea. Nu înţelesese explicaţia, dar ştia că tatăl ei are întotdeauna dreptate în lucrurile astea — sau, cel puţin, avusese *aproape* întotdeuna dreptate. Mă rog, fusese discuţia aceea despre oră în Noua Zeelandă, când mami a spus un lucru şi tati insista că nu e aşa. Pe urmă, trei ore

mai târziu, tati a coborât din birou şi a spus că-i pare rău şi că se înşelase, iar mami avusese dreptate tot timpul. A fost amuzant! „Pariez că şi mami putea fi savant dacă ar fi vrut" îşi spuse în sinea ei. „Şi n-ar fi avut dispoziţii din astea ciudate ca tata"

Jessica fu ceva mai atentă să pună următoarea întrebare în momentul în care tatăl său tocmai terminase ceea ce făcea şi nu începuse încă ceva nou. „Tati? Ştiu că bolovanul n-o să cadă, dar cum ar fi dacă ne-am imagina că ar cădea şi că am fi închişi aici pentru toată viaţa. S-ar face foarte întuneric? Am fi în stare să respirăm?

„Ce gânduri negre!" răspunse tatăl Jessicăi. Apoi privi atent bolovanul de deasupra intrării. „Hmm", spuse, „da, cred că bolovanul va acoperi intrarea aproape complet. Va rămâne cu siguranţă loc pentru ca aerul să intre şi să iasă, aşa că nu ne vom sufoca. Cât despre lumină, cred că va rămâne spărtura aia rotundă de deasupra care va lăsa să intre ceva lumină, dar va fi desigur foarte întuneric - mult mai întuneric decât acum. Însă sunt sigur că vom putea vedea destul de bine după ce ne obişnuim. Mă tem că n-ar fi prea frumos, totuşi. Dar pot să-ţi spun un lucru: dacă ar trebui să trăiesc aici tot restul vieţii cu cineva, aş prefera să fiu cu minunata mea Jessica decât cu oricine altcineva — şi cu mami, desigur.

Jessica îşi aminti de ce îl iubea pe tati aşa de tare. „Şi eu vreau ca mama să fie aici, pentru că în următoarea întrebare am să presupun că bolovanul a căzut înainte ca eu să mă fi născut, iar tu şi cu mami m-aţi crescut aici în peşteră de la început... şi am putut trăi pentru că am mâncat din plantele tale caraghioase."

Tati se uită puţin ciudat, dar nu spuse nimic.

„Atunci, n-aş fi văzut nici un fel de viaţă alta decât cea din peşteră. Cum aş fi putut şti ce e în lumea de afară? Aş fi putut şti că sunt copaci în ea, şi păsări, şi iepuri şi alte lucruri? Desigur, mi-ai fi putut *spune* toate astea, deoarece tu le-ai văzut înainte de a fi prins aici, dar cum aş fi putut *eu* şti - vreau să zic, cum aş fi putut şti eu însămi, fără să trebuiască să cred tot ce-mi spui tu?"

Tatăl se opri şi se gândi câteva minute. Apoi spuse: „Păi, presupun că, din când în când, în câte-o zi însorită, o pasăre ar fi putut zbura exact prin dreptul crăpăturii, astfel încât i-am fi putut vedea umbra pe peretele din fund. Desigur, umbra ar fi oarecum deformată pe peretele ăsta neregulat, dar am fi putut învăţa cum s-o corectăm. Dacă crăpătura ar fi mică şi destul de rotundă, atunci pasărea ar putea arunca o umbră destul de bine conturată; dacă nu, ar trebui să ne imaginăm de asemenea o serie de corecţii. Pe urmă, dacă aceeaşi pasăre ar zbura pe aici de mai multe ori, am putea obţine o imagine destul de bună a felului în care arată, în care zboară şi aşa mai departe, toate doar din umbra ei. Apoi, dacă soarele ar fi destul de jos pe cer, s-ar putea întâmpla ca un copac, situat chiar între soare şi

crăpătura noastră să apară ca o umbră pe perete, cu frunzele în vânt, astfel încât am putea obţine şi o imagine de copac. Şi poate că, din timp în timp, un iepure ar putea sări prin faţa crăpăturii astfel încât să ni-l putem imagina şi pe el tot din umbra de pe perete."

„Interesant", spuse Jessica. Tăcu câteva clipe, apoi spuse: „Crezi că ar fi fost posibil să faci vreo descoperire ştiinţifică reală, aici, în peşteră? Să ne imaginăm că am făcut o mare descoperire despre lumea de afară şi apoi am ţine una din acele conferinţe, încercând să-i convingem pe toţi ceilalţi că avem dreptate. Desigur, toţi oamenii de la conferinţă (ca şi tine) ar fi trebuit să fie crescuţi în peşteră, altfel ar fi vorba doar de o scamatorie. Dar e în regulă, ei ar fi putut să crească în peşteră dacă am fi avut tone de plante din astea ciudate şi toţi am fi trăit mâncând din ele!"

De data asta tatăl Jessicăi păru vizibil tulburat, dar tot nu spuse nimic. Pentru câteva minute tăcu, părând că se gândeşte adânc. Apoi spuse: „Da, cred că ar fi posibil. Dar, vezi tu, cel mai greu lucru ar fi să-i convingem că lumea din afară există. Tot ce vor şti ei despre ea vor fi doar umbre şi felul în care acestea se mişcă sau îşi modifică forma, din timp în timp. Pentru ei, umbrele complicate şi mişcătoare de pe zidul peşterii reprezintă tot ceea ce există în lume. Deci, o parte a conferinţei noastre va trebui să-i convingă mai întâi că există o lume din afară, cea la care teoria noastră se referă. De fapt, aceste două lucruri merg împreună. O bună teorie despre lumea din afară va fi o parte importantă a încercării de a-i convinge pe oameni că există ceva în afara peşterii!"

„Oh, tati, şi care este teoria noastră?"

„Nu te grăbi....stai puţin...uite-o: pământul se învârte în jurul soarelui!"

„*Asta* nu e o teorie prea nouă."

„Nu - e chiar veche de aproape douăzeci şi trei de secole - un timp aproape la fel de lung ca cel în care bolovanul a stat deasupra intrării. Dar, în povestea ta, noi toţi ne-am petrecut vieţile în peşteră şi nimeni n-a auzit de o astfel de idee până acum. Va trebui să-i convingem mai întâi că există un astfel de lucru ca *soarele* - sau ca pământul. Ideea ar fi ca simpla eleganţă cu care teoria noastră va explica toate tipurile de detalii ale mişcării luminii şi umbrelor să-i convingă în cele din urmă pe toţi participanţii la conferinţă nu numai că există acolo afară un lucru atât de strălucitor pe care-l numim Soare, dar şi că Pământul este într-o continuă mişcare de rotaţie în jurul lui, învârtindu-se în jurul axei proprii."

„Şi va fi foarte greu să-i convingem?"

„Cu siguranţă! De fapt, vom avea de făcut două feluri de lucruri foarte diferite. Mai întâi, va trebui să arătăm că teoria noastră simplă explică într-un mod foarte precis un imens număr de date detaliate privind felul în care pata de lumină

şi umbrele se mişcă de-a lungul peretelui. În acest punct, unii dintre oameni s-ar putea lăsa convinşi; alţii însă, vor spune că există o teorie mult mai plină de bun simţ, conform căreia soarele este cel care se învârte în jurul pământului. În detaliile ei, o astfel de teorie va sfârşi prin a fi mult mai complicată decât cea pe care o propunem noi, însă oamenii vor prefera să creadă în teoria complicată pur şi simplu pentru că vor refuza să creadă (cu oarecare dreptate) că peştera noastră se mişcă cu aproximativ o sută de mii de kilometri pe oră, cum ar cere-o teoria noastră."

„Chiar face asta?"

„Da, cam aşa ceva. Astfel, pentru a doua parte a demonstraţiei noastre, va trebui să ne schimbăm complet tactica şi să facem lucruri pe care cei mai mulţi le vor considera irelevante. Vom rostogoli bile de-a lugul unor planuri înclinate, vom pune în mişcare penduluri şi alte chestii de acest fel — doar pentru a arăta că legile fizice care guvernează comportamentul lucrurilor din peşteră ar fi neafectate de faptul că peştera se mişcă sau nu în jurul pământului cu o asemenea viteză. Acesta a fost unul dintre lucrurile importante pe care Galileo Galilei le-a avut de arătat — ţi-l aminteşti, cel din cartea pe care ţi-am dat-o."

„Sigur că-mi amintesc! Numai că toate lucrurile sună teribil de complicat. Pariez că o grămadă de oameni vor adormi la conferinţa noastră, exact aşa cum am văzut că adorm la conferinţele reale când le vorbeşti tu".

Tatăl Jessicăi roşi uşor. „Bănuiesc că ai dreptate! Numai că mi-e teamă că asta e ştiinţa: grămezi de detalii, multe dintre ele părând foarte plicticoase şi uneori complet irelevante pentru imaginea pe care încerci s-o creionezi, chiar dacă teoria finală are ceva surprinzător de simplu, ca ideea noastră că Pământul se învârteşte în jurul acelui obiect numit „Soare". Unii oameni ar spune că nu simt nevoia să-şi bată capul cu toate detaliile astea plicticoase, deoarece ideea li se pare în orice caz plauzibilă. Însă adevăraţii sceptici vor vrea să verifice totul, căutând posibile portiţe de scăpare.

„Mulţumesc, tati. Totdeauna mi-a plăcut să vorbeşti cu mine despre lucruri din astea, când devii din ce în ce mai entuziasmat şi mai roşu la faţă. Dar putem să ne întoarcem acum? Se face întuneric şi sunt obosită, flămândă — şi mi-e un pic frig."

„Hai, atunci." Tatăl Jessicăi îşi puse braţul în jurul umerilor ei şi ieşiră împreună din peştera care se întuneca. La ieşire, Jessica se uită din nou la bolovan.

„Ştii ceva, tati? Cred că sunt de acord cu tine. Bolovanul ăla o să stea acolo sus cel puţin încă douăzeci şi trei de secole!"

Partea întâi

DE CE AVEM NEVOIE DE O NOUĂ FIZICĂ PENTRU A ÎNȚELEGE MINTEA

Non-computabilitatea gândirii conștiente

1. CONŞTIINŢA ŞI CALCULUL

1.1. Conştiinţă şi ştiinţă

Care este scopul final al ştiinţei? Sunt atributele *materiale* ale Universului nostru singurele care pot fi studiate cu metode ştiinţifice? Trebuie existenţa noastră mentală să rămână definitiv în afara ştiinţei? Sau vom ajunge într-o zi la o înţelegere ştiinţifică reală a misterului plin de umbre al minţii noastre? Este fenomenul conştiinţei umane ceva situat deasupra scopului cercetării ştiinţifice, sau puterea metodei ştiinţifice va ajunge odată să rezolve problema existenţei sinelui conştient ?

Există specialişti care consideră că suntem aproape de înţelegerea ştiinţifică a conştiinţei; că acest fenomen nu conţine nici un mister şi că toate ingredientele pot fi deja puse la locul lor. Aceştia argumentează că doar extraordinara complexitate şi organizarea atât de sofisticată a creierului nostru sunt cele care, în prezent, limitează înţelegerea minţii umane — o complexitate şi o sofisticare ce nu trebuie desigur subestimate — şi că nu există nimic principial în plus faţă de rezultatele noastre ştiinţifice actuale. La extrema cealaltă sunt cei care susţin că problemele minţii şi spiritului — şi chiar misterul conştiinţei umane — sunt lucruri despre care nu putem spera vreodată să fie în mod adecvat abordate de procedurile reci şi calculate ale unei ştiinţe lipsite de sensibilitate.

În această carte voi încerca să privesc problema conştiinţei dintr-un punct de vedere ştiinţific. Dar voi susţine cu tărie — în baza unei demonstraţii

ştiinţifice — că din imaginea noastră ştiinţifică lipseşte un ingredient esenţial. Acest ingredient ar fi necesar pentru ca problema centrală a minţii umane să poată fi înglobată într-o descriere ştiinţifică a lumii. Voi susţine că acest ingredient nu este, în sine, ceva situat deasupra ştiinţei, dar că avem, fără îndoială, nevoie de o altă descriere ştiinţifică extinsă şi mai adecvată a lumii. În partea a doua a acestei cărţi voi încerca să-l conduc pe cititor într-o direcţie foarte precisă, care să ducă la o astfel de extindere a imaginii noastre actuale asupra Universului fizic. Este o direcţie care implică o schimbare importantă în legile de bază ale fizicii şi voi fi foarte explicit asupra naturii schimbărilor necesare şi asupra modului în care acestea se pot aplica biologiei creierelor noastre. Chiar cu înţelegerea actuală, limitată asupra naturii acestui ingredient — lipsă, putem începe să ghicim unde anume ar trebui el să lase urme şi cum ar trebui să contribue la tot ceea ce fundamentează trăirile noastre conştiente şi acţiunile noastre.

Cu toate că, în mod necesar, unele dintre argumentele pe care la voi da nu sunt simple, voi încerca să explic lucrurile cât mai clar, utilizând pe cât posibil numai noţiuni elementare. În anumite puncte voi introduce unele tehnici matematice, dar numai acolo unde sunt strict necesare sau sunt de ajutor pentru a mări claritatea discuţiei.

Deşi am învăţat să nu aştept ca cineva să se lase convins de argumentele pe care la voi prezenta, aş dori să sugerez că aceste argumente trebuie cercetate cu atenţie şi cu o consideraţie lipsită de pasiuni; căci ele construiesc un caz ce nu trebuie ignorat.

O imagine ştiinţifică asupra lumii care nu înglobează problemele conştiinţei nu poate avea pretenţii serioase de completitudine. Conştiinţa este parte a Universului nostru şi deci orice teorie fizică care nu-i face loc, eşuează în încercarea de a da o descriere fidelă a lumii. Trebuie să menţionez că nu există nici o teorie fizică, biologică sau vreo teorie a calculului, care să se apropie suficient de o explicare a conştiinţei şi, în consecinţă, a inteligenţei; dar asta nu ne împiedică să căutăm una. Cu o astfel de aspiraţie în minte sunt prezentate argumentele acestei cărţi.

Probabil, într-o zi, o întreagă colecţie de idei adecvate va fi adusă în discuţie. Dacă va fi aşa, perspectiva noastră filosofică se va modifica profund. Totuşi, orice cunoaştere ştiinţifică este o sabie cu două tăişuri. Ce facem de faţă cu cunostintele noastre stiintifice, asta-i o altă chestiune. Să încercăm să vedem

unde ne-ar putea duce punctele noastre de vedere asupra ştiinţei şi asupra minţii.

1.2. Pot roboţii salva această lume tulburată?

Dacă deschidem un ziar sau ne uităm la televizor, ni se pare că suntem asaltaţi continuu de roadele stupidităţii umane. Ţări sau părţi de ţări, se opun una alteia şi confruntările pot, din când în când, să degenereze în războaie hidoase. Fervoarea religioasă excesivă, naţionalismul, interesele etnice diferite, diferenţele lingvistice sau culturale, interesele egoiste ale unor demagogi pot avea ca rezultat continuarea agitaţiei şi a violenţei, canducând la atrocităţi de nedescris. Regimuri autoritare şi opresive încă mai subjugă popoare ţinându-le sub control prin folosirea torturii şi a plutonului de execuţie. Mai mult, aceia care sunt oprimaţi şi care ar trebui să aibă interese comune, sunt adeseori în conflict şi, când obţin libertatea, pe care o doresc de multă vreme, pare că aleg să utilizeze această libertate în moduri autodistructive. Chiar în acele ţări norocoase unde există prosperitate şi libertăţi democratice, resurse şi forţe umane sunt risipite într-un mod de neînţeles. Nu este acest lucru o indicaţie clară asupra stupidităţii generale a Omului? Deşi credem despre noi că reprezentăm apogeul inteligenţei în lumea animală, această inteligenţă nu pare suficientă pentru a controla problemele cu care societatea noastră ne confruntă tot timpul.

Totuşi, realizările pozitive ale inteligenţei noastre nu pot fi contestate. Printre aceste realizări se numără desigur impresionanta noastră ştiinţă şi tehnologie. Într-adevăr, deşi trebuie recunoscut că unele dintre fructele acestei tehnologii au o valoare discutabilă pe termen lung (sau scurt), aşa cum stau mărturie numeroasele probleme de mediu, sau teama de o catastrofă globală indusă tehnologic, aceeaşi tehnologie a fost cea care a creat societatea modernă, cu confortul ei, cu eliberarea de frică, boală şi sărăcie şi cu posibilităţile ei mari de dezvoltare intelectuală şi estetică, sau de comunicare globală. Dacă această tehnologie a deschis atâtea posibilităţi şi, într-un sens, ne-a ajutat la dezvoltarea scopurilor şi posibilităţilor propriului nostru eu fizic, n-am putea spera mai multe de la viitor?

Simţurile noastre au fost mult extinse de tehnologie atât în trecut cât şi în vremurile moderne. Vederea noastră a fost ajutată şi puterea ei a crescut enorm cu

ochelari, oglinzi, telescoape, microscoape de toate felurile şi, mai mult, cu video camere, televiziune şi altele. Auzul a fost ajutat mai întâi de cornetele auditive, iar acum de minuscule dispozitive electronice şi a fost mult extins de telefoane, comunicaţii radio şi sateliţi. Avem biciclete, trenuri, automobile, vapoare, aeroplane, să ne ajute şi să transceadă mijloacele noastre naturale de locomoţie. Memoria noastră este ajutată de cărţi tipărite, filme şi de uriaşele capacităţi de memorare ale *calculatoarelor electronice*. Capacitatea noastră de calcul simplă şi de rutină, sau extrem de savantă, este extinsă enorm de posibilităţile calculatoarelor moderne. Astfel, tehnologia ne furnizează nu numai o expansiune a posibilităţilor noastre fizice, dar şi o lărgire a capacităţilor noastre mentale, prin mărirea posibilităţilor de executare a sarcinilor de rutină. Ce se poate însă spune despre sarcinile care nu sunt de rutină — sarcini care necesită inteligenţă autentică ? Este firesc să ne întrebăm dacă acestea se vor putea adăuga posibilităţilor tehnologiei bazată pe calculatoare.

Pentru mine este clar că societatea tehnologică (bazată pe calculatoare) are, cel puţin într-o privinţă, un mare potenţial de dezvoltare a inteligenţei. Mă refer aici la posibilităţile educative ale societăţii noastre, societate care poate câştiga enorm din diferitele aspecte ale tehnologiei — dar numai dacă le utilizează cu sensibilitate şi înţelegere. Tehnologia oferă posibilităţi prin: producţia de cărţi, filme, televiziune, sisteme interactive de diferite tipuri, bazate pe calculatoare. Acestea oferă multe posibilităţi pentru a ne extinde capacităţile intelectuale sau pentru a le distruge. Mintea umană este capabilă de mult mai mult decât are de obicei şansa să realizeze. Din nefericire, posibilităţile menţionate sunt prea adesea risipite, iar minţile celor tineri (sau batrâni) nu primesc în nici un fel deschiderile pe care, fără îndoială, le-ar merita. Mulţi cititori se vor întreba, dacă nu există o altă posibilitate pentru dezvoltarea capacităţilor mentale şi anume acea străină 'inteligenţă' electronică care tocmai începe să răsară din dezvoltarea extraordinară a tehnologiei calculatoarelor. Într-adevăr, adesea ne întoarcem spre calculatoare pentru asistenţă intelectuală. Sunt multe împrejurări în care inteligenţa umană neajutată este departe de a prevedea consecinţele probabile ale unor acţiuni alternative. Astfel de consecinţe se află mult deasupra posibilităţilor de calcul ale minţii umane; de aceea este de aşteptat ca în viitor calculatoarele să-şi extindă acest rol, acolo unde rezultatele unor calcule prea complicate pot aduce servicii de nepreţuit inteligenţei umane.

Vor putea calculatoarele să ofere şi mai mult? Mulţi experţi susţin că , cel puţin în principiu, calculatoarele oferă potenţialul pentru o *inteligenţă artificială* care, în cele din urmă, o va depăşi pe a noastră [1]. Când roboţii comandaţi prin calculator vor atinge nivelul 'echivalenţei umane' , nu va mai dura mult până când vor lăsa cu mult în urmă nivelul nostru modest. Doar atunci, pretind aceşti experţi, vom avea o autoritate cu suficientă inteligenţă, înţelepciune şi înţelegere, încât să fie capabilă să rezolve dificultăţile acestei lumi pe care umanitatea le-a creat.

Cât va dura până când această stare fericită va sosi ? Aici nu există un consens între experţi. Unii măsoară acest interval în secole, în vreme ce alţii susţin că echivalenţa umană este la distanţă doar de câteva decade. Aceştia din urmă îşi bazează opiniile pe viteza exponenţială de creştere a puterii calculatoarelor şi pe comparaţia ce o fac între viteza şi precizia tranzistorilor şi viteza relativ mică şi imprecizia acţiunii neuronilor. Într-adevăr, circuitele electronice sunt acum cam de un milion de ori mai rapide decât viteza de răspuns a neuronilor din creier (viteza fiind cam 10^9/s pentru tranzistori şi numai de aproximativ 10^3/s pentru neuroni)[1] şi ei au o precizie mare şi o comportare constantă în timp, pe care neuronii nu o au în nici un caz. Mai mult, comportarea creierului este într-o mare măsură aleatoare şi pare că ar putea fi enorm îmbunătăţită prin organizarea precisă a circuitelor electronice imprimate.

Există unele zone în care structura neuronală a creierului are un avantaj numeric asupra calculatoarelor actuale, dar acest avantaj numeric este de scurtă durată. Se argumentează că, faţă de numărul total de neuroni (câteva sute de mii de milioane), creierul uman depăşeşte, momentan, calculatoarele în raport cu numărul lor de tranzistori. Mai mult, în medie, există un număr mult mai mare de *legături* între diferiţi neuroni, comparativ cu legăturile între tranzistorii unui calculator. În particular, celulele lui Purkinje, din cerebel, pot avea până la 80000 de terminaţii sinaptice (legături între neuroni), în vreme ce pentru calculatoare numărul corespunzător este de trei sau patru. (Voi face câteva comentarii despre cerebel, mai târziu, în 1.14 şi 8.6). În plus, în calculatoarele de astăzi, tranzistorii sunt distribuiţi majoritar în circuitele memoriei (în defavoarea circuitelor de calcul, din microprocesor), în contrast cu situaţia din creier unde acţiunea de calcul pare

[1] *Chipul Intel Pentium* are peste 3 milioane de tranzistori pe o "felie de silicon" de mărimea unei unghii, fiecare capabil sa opereze 113 milioane de instrucţiuni pe secundă.

a fi mult mai distribuită.

Aceste avantaje temporare ale creierului pot fi uşor depăşite în viitor când sistemele paralele de calcul vor fi mai dezvoltate. Este un mare avantaj al calculatoarelor că diferite unităţi pot fi combinate pentru a forma unităţi tot mai mari, astfel încât numărul total al tranzistorilor poate, în principiu, să crească fără limite. În plus, există, în aşteptare, o revoluţie tehnologică prin care tranzistorii şi conexiunile (cablajele) din calculatorele actuale să fie înlocuite cu dispozitive laser care, probabil, vor mări în mod spectaculos viteza, puterea şi miniaturizarea. Fundamental, creierul nostru va părea tot mai mărginit, cu numărul său de celule pe care le are în prezent şi cu constrângerile suplimentare (ca aceea că trebuie să crească pornind de la o singură celulă). Dimpotrivă, calculatoarele pot fi în mod deliberat construite astfel, încât să satisfacă tot ceea ce se cere. Deşi în continuare voi prezenta factorii importanţi, pe care momentan nu-i iau în considerare în această prezentare (în special un anumit nivel semnificativ de activitate care stă la baza activităţii neuronului), totuşi este impresionant faptul că, deşi momentan calculatoarele nu întrec creierul, totuşi ele o vor face nu peste multă vreme.

Astfel, dacă ar fi să credem cele mai tari argumente ale propovăduitorilor entuziaşti ai inteligenţei artificiale şi am accepta că roboţii comandaţi de calculatoare vor putea eventual — şi încă nu peste mult timp — să depăşească toate capacităţile umane, atunci calculatoarele vor putea face cu mult mai mult decât ceea ce poate face inteligenţa noastră. Ele vor avea de fapt o inteligenţă imensă, a lor, proprie. Vom putea *atunci* să apelăm la această inteligenţă superioară, pentru sfaturi şi expertiză în orice domeniu — şi tulburările provocate de oameni în lume vor putea fi, în fine, rezolvate.

Dar apar alte consecinţe logice ale acestei dezvoltări potenţiale care ar putea cu adevărat să ne alarmeze. Nu vor face aceste calculatoare ca însăşi specia umană să devină inutilă ? Dacă se va constata că roboţii ghidaţi de calculatoare vor fi superiori în toate privinţele, atunci nu vor ajunge ei la concluzia că pot conduce lumea mai bine, fără ajutorul nostru ? Umanitatea însăşi va deveni perimată.

Poate, dacă suntem norocoşi, roboţii de mâine ne vor păstra ca pe nişte animale răsfăţate, aşa cum spunea odată Edward Fredklin; sau, dacă suntem destul de deştepţi, am putea transfera roboţilor 'structurile noastre proprii de informaţie', aşa cum insista Hans Moravec (1988).

Dar poate că nu vom fi nici atât de norocoşi şi nici atât de deştepţi

1.3. *A, B, C, D*-ul calculului şi gândirii conştiente

Este depăşirea capacităţii umane de către calculator doar o chestiune de putere de calcul, de viteză, precizie, memorie, sau poate, de modul de interconectare? Pe de altă parte, putem face oare ceva (cu creierul nostru) care să nu poată fi descris în termeni computaţionali? Cum ar putea sentimentele care implică conştiinţa conştientă[2] – fericire, durere, dragoste, sensibilitate estetică, voinţă, înţelegere etc., să se potrivească într-un astfel de tablou computaţional ? Vor avea cu adevărat calculatoarele viitorului minte? Are prezenţa gândirii conştiente influenţă, în vreun fel, asupra comportamentului? Se justifică să vorbim despre astfel de lucruri în termeni ştiinţifici, sau ştiinţa nu este competentă să atace subiecte legate de conştiinţa umană?

Mi se pare că există cel puţin patru puncte de vedere care pot fi identificate:

A Toată gândirea este calcul; chiar trăirile care implică conştiinţa conştientă sunt produse prin executarea unor calcule.

B - Conştienţa este o caracteristică a acţiunilor fizice ale creierului; dar orice acţiune fizică poate fi simulată computaţional. Pe de altă parte, simularea computaţională nu poate, în sine, trezi conştienţa.

C Acţiunea fizică corespunzătoare a creierului provoacă conştienţa, dar nici măcar această acţiune fizică nu poate fi simulată computaţional în mod corespunzător.

D Conştienţa nu poate fi explicată în termeni fizici, nici în termeni computaţionali şi în nici un fel de alţi termeni ştiinţifici.

Punctul de vedere exprimat în *D* care neagă posibilitatea de explicare

[2] Pentru diferenţa dintre conştienţă – *awareness* şi conştiinţă – *consciousness* vezi paragraful 1.12 (pag.60)

fizică şi priveşte mentalul ca pe ceva inexplicabil în termeni ştiinţifici, este punctul de vedere mistic; cel puţin unele aspecte din $\mathcal{D}$ par implicate în acceptarea doctrinelor religioase. Poziţia mea personală este că aceste probleme ale minţii, chiar dacă se plasează inconfortabil în raport cu înţelegerea ştiinţifică din zilele noastre, nu trebuie privite ca elemente în afara domeniului ştiinţei. Dacă ştiinţa nu este capabilă încă să spună mai mult despre problemele gândirii, atunci cu siguranţă ştiinţa trebuie să-şi lărgească scopurile pentru a se adapta acestor probleme, modificându-şi, la nevoie, metodele. Respingând misticismul datorită negării existenţei criteriilor ştiinţifice pentru progresul cunoaşterii, cred, în acelaşi timp, că, într-o ştiinţă şi o matematică extinse, se va găsi suficient mister pentru a cuprinde chiar şi misterul minţii. Voi dezvolta câteva din aceste idei în cele ce urmează dar, pentru moment, trebuie să spun că resping punctul $\mathcal{D}$; şi mă voi strădui să înaintez pe drumul pe care ştiinţa ni l-a deschis tuturor. Dacă cititorul acceptă punctul de vedere $\mathcal{D}$, îl invit să vadă cât se poate avansa pe drumul ştiinţei, încercând să înţeleagă unde sunt eu convins că ne duce, în cele din urmă, acest drum.

Să considerăm acum cealaltă extremă, punctul de vedere $\mathcal{A}$. Aceia care aderă la acest punct de vedere sunt susţinătorii a ceea ce se numeşte, de obicei, inteligenţă artificială tare (**IA** *tare* sau *strong Artificial Intelligence*), sau funcţionalism [4] (cu toate că unii folosesc termenul funcţionalism într-un fel care ar putea include de asemenea anumite versiuni ale punctului de vedere $\mathcal{C}$). $\mathcal{A}$ este privit de unii ca singurul punct de vedere acceptabil ca atitudine ştiinţifică. Alţii sunt de părere că, $\mathcal{A}$ este o absurditate căreia nu trebuie să i se dea atenţie. Fără îndoială că există o mulţime de versiuni ale acestui punct de vedere ($\mathcal{A}$). (Vezi Sloman (1992) pentru o listă lungă de variante computaţionale.) Diferenţa dintre aceste puncte de vedere priveşte modul în care este definit calculul şi desfăşurarea lui. Într-adevăr, există aderenţi ai lui $\mathcal{A}$ care susţin că nu sunt suporteri ai concepţiei **IA** tare, deoarece interpretează termenul 'calcul' altfel decât în concepţia **IA** convenţională (Edelman 1992). Vom detalia acest subiect al computabilităţii în paragraful 1.4 Pentru moment este suficient să acceptăm că acest termen, *computabilitate*, înseamnă ceea ce un calculator obişnuit este capabil să facă în mod normal.

Alţi susţinători ai lui $\mathcal{A}$ se deosebesc prin interpretarea ce o dau cuvintelor 'conştienţă' şi 'conştiinţă'. Unii nici măcar nu acceptă că poate exista

un astfel de fenomen, al 'conştiinţei conştiente', în vreme ce alţii acceptă existenţa acestuia privind-o ca un fel de proprietate emergentă (vezi paragrafele 4.3 şi 4.4), proprietate ce apare ori de câte ori intervine un nivel de dificultate (sau sofisticare sau autoreferenţiere) suficient de mare, în calculul care trebuie executat. Voi prezenta propria mea interpretare a termenilor 'conştiinţă' şi 'conştienţă' în paragraful 1.12.

Pentru moment, posibilele diferenţe de interpretare nu sunt de prea mare importanţă.

Punctul de vedere $\mathcal{A}$ (concepţia **IA** tare) este cel împotriva căruia se îndreaptă cu precădere argumentele mele din cartea precedentă, *Mintea noastră ... cea de toate zilele*[3] Până şi dimensiunile cărţii arată că, deşi eu nu cred că $\mathcal{A}$ este corectă, o privesc ca pe o posibilitate serioasă căreia trebuie să i se acorde o atenţie deosebită. $\mathcal{A}$ este rezultatul unei mari încrederi în ştiinţă: lumea fizică este considerată a opera complet computaţional. La una din extremele acestui punct de vedere, Universul însuşi este considerat a fi un gigantic calculator [5]; anumite secvenţe de calcul pe care acesta le execută provoacă stările conştienţei constitutive minţii noastre.

Eu presupun că acest punct de vedere — că sistemele fizice trebuie privite în principal ca entităţi computaţionale — derivă în parte din rolul tot mai mare pe care simularea computaţională îl joacă în ştiinţa secolului XX şi, în parte, din credinţa că obiectele fizice sunt în principal structuri informaţionale (într-un anumit sens) şi că se supun legilor calculabilităţii matematice. Cea mai mare parte a celulelor corpului şi creierului uman se înlocuiesc permanent şi numai *pattern*-ul persistă. Mai mult, materia pare să aibă o existenţă tranzientă deoarece ea poate fi convertită dintr-o formă în alta. Chiar şi masa unui corp, având o măsură precisă a cantităţii de materie, poate fi convertită în energie pură (conform celebrei formule a lui Einstein $E=mc^2$), aşa încât chiar substanţa pare a fi capabilă a se transforma în ceva care aparţine matematicii teoretice. Mai mult, mecanica cuantică ne spune

---

[3] Roger Penrose 'The Emperor's New Mind. Concerning Computers, Minds, and the Laws of Physics', Oxford University Press, 1989, ISBN 0-19-851973-7, tradusă în limba română de Cornelia C.Rusu şi Mircea V.Rusu, cu titlul "Mintea noastră ... cea de toate zilele : Despre gândire, fizică şi calculatoare", Editura Tehnică, Bucureşti 1996, ISBN 973-31-0969-X În continuare, toate referirile autorului la pasaje din 'The Emperor's New Mind' vor fi transpuse relativ la această versiune în limba română folosind prescurtarea MN (n.t.)

că particulele materiale sunt cu precădere 'unde de informaţie' (vom examina mai în detaliu subiectul în partea a doua). Prin urmare, materia este nebuloasă şi tranzientă; şi nu este deloc lipsit de sens să presupunem că persistenţa sinelui ar putea avea mai multe în comun cu conservarea formelor decât cu actualele particule materiale.

Chiar dacă nu credem că e potrivit să privim Universul ca pe un simplu calculator, ne-am putea simţi tentaţi de punctul de vedere $\mathcal{A}$. Să presupunem că avem un robot comandat de un calculator şi care răspunde la întrebări exact ca o fiinţă umană. Îl întrebăm cum se simte şi constatăm că răspunde în conformitate cu 'sentimentele' pe care de fapt le posedă. Ne spune că este conştient că-i fericit sau trist, că poate percepe culoarea roşie şi că se teme de întrebările despre 'minte' şi 'sine'. El poate chiar să dea expresie unui paradox privind necesitatea acceptării faptului că alte fiinţe (în special fiinţe umane) să fie privite ca posedând o conştiinţă similară cu cea pe care spune că simte că o are. De ce ar trebui să nu credem pretenţiile sale de a fi conştient, de a se minuna, de a fi bucuros sau de a simţi durerea, atunci când ştim că avem prea puţine certitudini cu privire la alte fiinţe umane, despre care acceptăm să le privim ca entităţi conştiente?

Argumentul operaţional pare să aibă o forţă considerabilă, după părerea mea, chiar dacă nu este pe de-a întregul decisiv. Dacă toate manifestările externe ale unui creier conştient, incluzând răspunsuri la noi şi noi întrebări, pot fi imitate cu succes de un sistem comandat de un calculator, atunci ar trebui să acceptăm că şi manifestările interne — conştiinţa însăşi — ar fi prezente în respectiva simulare.

Acceptarea acestui tip de argument — care se numeşte test Turing [6] — este esenţa diferenţei dintre $\mathcal{A}$ şi $\mathcal{B}$, În conformitate cu $\mathcal{A}$, un robot comandat de un calculator, care poate răspunde în mod susţinut la întrebări şi se comportă de parcă ar avea conştiinţă, trebuie considerat că are cu adevărat conştiinţă — în vreme ce, în conformitate cu $\mathcal{B}$, un robot s-ar putea comporta exact ca o persoană conştientă fără ca asta să însemne că are conştiinţă. Atât $\mathcal{A}$, cât şi $\mathcal{B}$, acceptă că un robot comandat de un calculator se poate comporta în mod convingător ca o persoană conştientă. Punctul de vedere $\mathcal{C}$, pe de altă parte, nu ar putea admite nici măcar posibilitatea unei simulări efective a conştienţei de către un robot controlat de un calculator. $\mathcal{C}$ spune că, în mod necesar, lipsa de conştienţă a unui robot trebuie să iasă la iveală în cele din urmă, după un interogatoriu destul de lung. Într-adevăr, $\mathcal{C}$ este mai *operaţional* decât $\mathcal{B}$ şi seamănă, în această privinţă,

mai mult cu $\mathcal{A}$ decât cu $\mathcal{B}$.

Ce se poate spune atunci despre $\mathcal{B}$? Cred că acesta poate fi privit ca punctul de vedere al 'bunului simţ ştiinţific', numindu-se uneori punctul de vedere **IA** *slab* (al inteligenţei artificiale slabe, *weak* AI). Ca şi $\mathcal{A}$, acesta susţine că toate obiectele fizice (ale acestei lumi) trebuie să se comporte în conformitate cu o teorie, care le permite să fie simulate computaţional. Pe de altă parte, respinge cu tărie părerea operaţională că lucrurile care se comportă exact ca o fiinţă conştientă trebuie să fie ele însele conştiente. Aşa cum a subliniat filozoful John Searle [7], simularea computaţională a unui proces fizic este un lucru complet diferit de procesul însuşi (simularea unui uragan, cu siguranţă, nu este un uragan!). În conformitate cu $\mathcal{B}$, prezenţa sau absenţa conştiinţei ar depinde foarte mult de care obiect fizic este 'responsabil cu gândirea' şi care anume sunt acţiunile fizice particulare pe care acest obiect le efectuează. Considerarea calculelor particulare, care se întâmplă să fie implicate în aceste acţiuni, este o chestiune secundară. Astfel, acţiunea unui creier biologic ar putea invoca conştiinţa, pe când simularea electronică corespunzătoare n-o va face. Nu este necesar, din punctul de vedere $\mathcal{B}$, ca distincţia să se facă între biologie şi fizică. Dar componenta materială a obiectului în discuţie (să spunem creier) este privită ca esenţială, nu doar acţiunea sa computaţională.

Punctul de vedere $\mathcal{C}$ este cel pe care, personal, îl consider ca fiind cel mai aproape de adevăr. Este mai operaţional decât $\mathcal{B}$, deoarece susţine că există manifestări externe ale obiectelor conştiente (să spunem creiere), care diferă de manifestările externe ale unui calculator: efectele externe ale conştiinţei nu pot fi simulate computaţional în mod adecvat. Voi oferi, la momentul potrivit, motivele care susţin această convingere. Deoarece $\mathcal{C}$, ca şi $\mathcal{B}$, susţine punctul de vedere fizicalist, după care mintea apare ca manifestare a comportamentului unui obiect fizic (creierul, dar poate nu numai creierul), rezultă că una dintre implicaţiile lui $\mathcal{C}$ este că nu orice acţiune fizică poate fi simulată computaţional în mod adecvat.

Dar cunoaşte oare fizica zilelor noastre vreo acţiune care să nu poată fi simulată în principiu de un calculator? Răspunsul nu-mi este complet limpede, dacă ceea ce se cere este o demonstraţie matematică. Ştim prea puţin despre acest subiect [8] pentru a putea formula cu precizie o teoremă matematică. Cu toate acestea, convingerea mea fermă este că astfel de acţiuni non-computaţionale ar trebui găsite într-un domeniu al fizicii ce se află în afara legilor actualmente

cunoscute ale fizicii.

Mai târziu, în această carte, voi reveni asupra motivelor, care provin chiar din fizică, motive ce mă determină să cred că o nouă înţelegere este cu adevărat necesară, într-o zonă care se află plasată intermediar între microcosmos (zona în care guvernează legile mecanicii cuantice) şi nivelul fizicii clasice. Totuşi, nu este unanim acceptat, printre fizicienii din zilele noastre, că o astfel de teorie ar fi necesară.

Astfel, există cel puţin două opinii diferite care pot fi circumscrise punctului de vedere $\mathcal{C}$ Unii susţinători ai lui $\mathcal{C}$ vor pretinde că înţelegerea prezentă a fizicii este perfect adecvată şi că ar trebui să căutăm tipuri mai subtile (sofisticate) de comportament în interiorul teoriei convenţionale, capabile să ne ducă în afara limitelor a ceea ce poate fi obţinut computaţional (de exemplu, dintre cele examinate în continuare, comportarea haotică (1.7), subtilităţi ale continuului — ca opus interacţiunilor discrete (1.8), probabilitatea cuantică). Pe de altă parte, există alţii care ar putea argumenta că fizica de astăzi nu ne oferă motive întemeiate pentru necalculabilitatea de tipul cerut. Pe parcurs, voi arăta care sunt argumentele esenţiale în sprijinul lui $\mathcal{C}$, în conformitate cu acest punct de vedere mai *tare*, care ar cere introducerea unei fizici fundamental modificate.

Unii consideră că toate acestea mă plasează pe mine în domeniul lui $\mathcal{D}$, deoarece susţin că trebuie să ne uităm dincolo de marginile ştiinţei cunoscute dacă vrem ca, vreodată, să găsim o explicaţie pentru fenomenul conştiinţei.

Însă există o diferenţă esenţială între această versiune *tare* a lui $\mathcal{C}$ şi punctul de vedere $\mathcal{D}$, în ceea ce priveşte metodologia. În conformitate cu $\mathcal{C}$, problema conştiinţei conştiente este o problemă ştiinţifică, chiar dacă o teorie corespunzătoare încă nu există. Eu susţin asta cu tărie. Cred că răspunsul trebuie căutat prin metode ştiinţifice, chiar dacă aceste metode vor fi obţinute prin extinderi pe care cu greu le putem imagina în prezent. Aceasta este diferenţa esenţială între $\mathcal{C}$ şi $\mathcal{D}$ oricât ar părea ele de asemănătoare în privinţa posibilităţilor şi realizărilor ştiinţei prezentului.

Poziţiile $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ şi $\mathcal{D}$, aşa cum au fost ele definite mai sus, reprezintă punctele extreme ale diferitelor atitudini care pot fi alese. Trebuie să accept că anumite persoane ar putea să simtă că punctul lor de vedere nu se potriveşte cu nici una dintre aceste categorii ci este situat undeva între ele, sau se intersectează cu unele din ele.

Există, cu siguranţă, multe gradaţii ale convingerilor între $\mathcal{A}$ şi $\mathcal{B}$ (vezi Sloman 1992). Există de asemeni un punct de vedere, adeseori exprimat, care poate fi considerat o combinaţie dintre $\mathcal{A}$ şi $\mathcal{B}$ (sau poate, mai bine, între $\mathcal{B}$ şi $\mathcal{D}$ — o posibilitate care se potriveşte de fapt cu dezbaterile noastre ce urmează. Conform acestei poziţii, activitatea creierului este, cu adevărat, cea a unui calculator, dar a unuia de o astfel de complexitate încât imitarea lui este deasupra capacităţii omului şi ştiinţei, fiind necesară creaţia divină a lui Dumnezeu — „cel mai bun programator în branşă"! [9]

1.4. Fizicalism şi mentalism

Trebuie să încep printr-o scurtă remarcă asupra felului în care termenii *fizicalism* şi *mentalism* sunt adeseori folosiţi pentru a descrie puncte de vedere opuse în conexiune cu problemele ridicate de $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ şi $\mathcal{D}$.

Deoarece $\mathcal{D}$ reprezintă o negare totală a *fizicalismului*, cei care cred în acest punct de vedere pot fi cu certitudine clasificaţi ca adepţi ai *mentalismului*. Cu toate acestea, nu-mi este deloc clar unde se poate trage linia dintre fizicalism şi mentalism, în relaţie cu celelalte trei puncte de vedere, $\mathcal{A}$ $\mathcal{B}$ şi $\mathcal{C}$.

Cred că susţinătorii lui $\mathcal{A}$ pot fi consideraţi drept adepţi ai fizicalismului şi sunt sigur că marea lor majoritate va adera la această clasificare. Totuşi, în spatele acestei împărţiri se ascunde un paradox. După cei care susţin punctul de vedere $\mathcal{A}$ construcţia *materială* a 'aparatului de gândire' este considerată nerelevantă. Doar calculul pe care această structură materială îl efectuează îi determină toate atributele. Cât despre calcule, acestea sunt elemente de matematică abstractă, separate de orice asociere cu obiecte materiale particulare. Astfel, după $\mathcal{A}$, atributele mentale sunt ele însele lucruri fără legătură cu obiectele fizice şi deci termenul de *fizicalism* ajunge să pară oarecum nepotrivit.

Punctele de vedere $\mathcal{B}$ şi $\mathcal{C}$, pe de altă parte, cer alcătuirii fizice a unui obiect să joace un rol esenţial în determinarea prezenţei posibile a 'mentalului'. În consecinţă, se poate foarte bine susţine că, mai degrabă aceste puncte de vedere şi nu $\mathcal{A}$, reprezintă posibilele atitudini fizicaliste. Totuşi, pare că o astfel de terminologie ar fi într-o oarecare contradicţie cu folosirea uzuală a cuvintelor, unde termenul 'mentalism' este mai degrabă considerat adecvat pentru $\mathcal{B}$ şi $\mathcal{C}$,

deoarece aici, calităţile mentale sunt privite ca fiind 'lucruri reale' şi nu doar epifenomene ce apar accidental atunci când anumite tipuri de calcul sunt efectuate.

Pentru a evita confuziile de mai sus, voi încerca să nu apelez la termenii *fizicalism* şi *mentalism* în discuţiile care urmează şi să mă refer, în schimb, la punctele de vedere specifice, $\mathcal{A}$ $\mathcal{B}$, $\mathcal{C}$ şi $\mathcal{D}$, definite anterior.

1.5. Calcul: proceduri top-down şi bottom-up

Până acum nu am explicat ce înţeles am dat termenului *calcul* atunci când l-am folosit în definiţiile $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ şi $\mathcal{D}$ din 1.3. Ce este *calculul*? Pe scurt, am putea înţelege acest termen ca denumind activitatea unui calculator obişnuit. Mai precis, ar trebui să luăm termenul într-un sens mai abstract: un *calcul* este acţiunea unei maşini Turing.

Dar ce este o maşină Turing? Un calculator idealizat (înaintaşul teoretic al calculatorului obişnuit din zilele noastre) care nu face greşeli, care poate funcţiona un timp infinit de lung şi care are o capacitate de memorie nelimitată. Voi fi mai explicit asupra modului în care maşinile Turing pot fi definite în paragraful 2.1 şi în Anexa A. (Pentru o mai detaliată introducere, cititorul interesat poate citi capitolul 2 din MN, p. 39-85, sau, de exemplu, Kleene (1952), Devis (1978) etc.)

Termenul *algoritm* este folosit în mod frecvent pentru descrierea unei maşini Turing. Consider, în cele ce urmează, termenul *algoritm* complet sinonim cu *calcul*. Acest lucru necesită o precizare, deoarece unii au un punct de vedere mai restrictiv asupra noţiunii de algoritm în comparaţie cu varianta propusă aici şi anume în sensul în care folosim aici noţiunea de *algoritm top-down*. Să încercăm deci, mai întâi, să vedem ce se înţelege prin termenii *top-down* şi *bottom-up* în contextul calculului.

O procedură de calcul se zice că este organizată *top-down* dacă a fost construită în concordanţă cu o procedură de calcul fixă, bine definită şi înţeleasă în mod clar, procedură care furnizează o soluţie clar precizată pentru o anumită problemă considerată. (Algoritmul lui Euclid de găsire a celui mai mare divizor comun a două numere naturale, aşa cum este descris în MN p.41, este un exemplu

simplu de algoritm *top-down*.) Prin contrast, în organizarea *bottom-up*[4], nu se specifică dintru început regulile de operare şi cunoştinţele utilizate şi, în loc de aceasta, procedura permite ca sistemul să 'înveţe' şi să-şi îmbunătăţească performanţele pe baza 'experienţei' Astfel, într-un sistem *bottom-up* regulile de operare pot fi supuse unor modificări continue. Trebuie să se permită sistemului să funcţioneze de multe ori, să-şi execute acţiunea asupra unor date de intrare furnizate continuu. După fiecare execuţie se face o verificare — posibil chiar de către sistemul însuşi — care modifică operaţiile, în lumina rezultatelor verificării, cu scopul de a îmbunătăţi calitatea rezultatelor. De exemplu, datele de intrare în sistem ar putea fi nişte fotografii reprezentând feţe umane, discretizate în mod corespunzător, iar sarcina sistemului ar fi să stabilească care dintre fotografii reprezintă aceeaşi persoană şi care nu. După fiecare rulare, performanţele sistemului sunt comparate cu răspunsurile corecte. Regulile de operare sunt atunci modificate în aşa fel, încât să conducă, la următoarea rulare, la o probabilă îmbunătăţire a performanţelor sistemului.

Detaliile despre cum se realizează o astfel de îmbunătăţire pentru un sistem particular *bottom-up*, nu sunt cu adevărat importante pentru noi aici. Există o mulţime de scheme diferite. Printre cele mai bine cunoscute sisteme *bottom-up* sunt aşa numitele *reţele neurale artificiale* (uneori denumite, în mod nefericit, pur şi simplu 'reţele neuronale' sau 'reţele neurale'[5]) care sunt programe ale unui calculator care învaţă (sau ale altui dispozitiv electronic special construit) bazate pe anumite idei despre organizarea sistemului de conexiuni dintre neuronii din creier, care să se perfecţioneze pe măsură ce sistemul câştigă experienţă (problema modului în care sistemul de interconexiuni dintre neuronii din creier se modifică, va fi foarte importantă pentru noi, în cele ce urmează; conform cu paragrafele 7.4 şi 7.7.). Evident, pot să existe sisteme de calcul care combină organizarea *top-down* cu cea *bottom-up*.

Important pentru scopurile noastre este faptul că, atât procedurile bottom-

---

[4] Definiţiile sunt importante întrucât termenii aceştia (*top-down*, *bottom-up*) au alte semnificaţii în alt context (n.t.)

[5] De la apariţia cărţii, domeniul acesta al *reţelelelor neuronale* s-a dezvoltat aproape exploziv. Termenii contestaţi aici de autor impunându-se în limba română, vom folosi în cele ce urmează această terminologie deja acceptată (n.t.)

up cât şi cele *top-down*, pot fi executate cu un calculator şi deci ambele pot fi incluse în ceea ce am desemnat drept calculabil sau algoritmic. Astfel, în cazul procedurilor *bottom-up*, mijlocul prin care sistemul îşi modifică procedura este el însuşi furnizat pe de-a-ntregul de o procedură calculabilă specificată în prealabil. De aceea, întregul sistem poate fi implementat pe un calculator obişnuit.

Deosebirea esenţială dintre un sistem *bottom-up* şi unul *top-down* constă în faptul că, în sistemele *bottom-up*, trebuie să existe posibilitatea ca experienţa trecută să fie memorată, iar lucrurile păstrate în memorie să poată fi încorporate în viitoarele acţiuni. Pentru moment nu vom intra în detalii, dar discuţia va fi reluată în paragraful 3.11.

Inteligenţa artificială (prescurtat **IA**) urmăreşte să imite comportarea inteligentă, la toate nivelurile, prin mijloace computaţionale. În acest scop s-au utilizat frecvent atât proceduri *top-down* cât şi proceduri *bottom-up*. Iniţial, părea că procedurile *top-down* sunt mai promiţătoare, dar, mai nou, sunt foarte răspândite sistemele de tip reţele neuronale *bottom-up*. Probabil că o anumită combinaţie de organizare *top-down* şi *bottom-up* va duce la sisteme care să fie în viitor de success în **IA**. Pot fi obţinute avantaje din fiecare tip de organizare. Organizarea *top-down* are succes în domeniile în care datele şi regulile de calcul sunt bine şi clar definite, ca în cazul unor probleme de matematică, în cazul sistemelor de calcul specializate în jocul de şah, sau în cazul sistemelor de diagnostic medical, în care mulţimea regulilor de diagnosticare a unor anumite maladii se bazează pe metode medicale general acceptate. Organizarea *bottom-up* este mai utilă atunci când criteriile de decizie nu sunt foarte bine precizate sau sunt incomplet înţelese, aşa cum se întâmplă la recunoaşterea feţelor sau a sunetelor, sau la prospectarea zăcămintelor minerale, unde îmbunătăţirea performanţelor pe baza experienţei este criteriul de bază în aprecierea comportării. În multe din aceste cazuri intervin atât organizarea *top-down* cât şi cea *bottom-up* (aşa se întâmplă cu sistemele de calcul pentru jocul de şah, care învaţă şi din experienţă, cât şi în cazul dispozitivelor de calcul folosite în prospecţiunile geologice, când anumite cunoştinţe precise de teorie geologică sunt încorporate în dispozitivul de calcul).

Cred că este cinstit să recunoaştem că, de fapt, calculatoarele şi-au putut demonstra o superioritate semnificativă asupra oamenilor, tocmai utilizând o organizare de tip *top-down*. Exemplul cel mai evident este calculul numeric, unde calculatoarele câştigă de departe competiţia cu oamenii şi, de asemenea în jocurile

'computaţionale'[6] cum sunt şahul sau jocul de dame, unde prea puţini jucători pot bate cele mai bune maşini (mai mult despre aceste lucruri în paragrafele 1.5 şi 8.2). Cu o organizare de tip *bottom-up* (reţele neuronale) calculatoarele pot atinge nivelul unor oameni obişnuiţi, dar bine antrenaţi, doar în câteva situaţii foarte limitate.

O altă clasificare a sistemelor de calcul se poate face diferenţiind arhitectura serială de cea paralelă. O maşină serială execută calculele unele după altele acţionând pas cu pas, în vreme ce calculatoarele paralele execută multe calcule independente simultan, rezultatele lor fiind puse împreună doar atunci când toate rezultatele parţiale au fost obţinute. În acest caz, studiile despre felul în care s-ar putea să opereze creierul au inspirat dezvoltarea unor sisteme paralele de calcul. Totuşi, se cuvine să precizăm că, de fapt, nu există o diferenţă de *principiu* între maşinile seriale şi cele paralele. Există întotdeauna posibilitatea de a simula o acţiune paralelă printr-una serială, chiar dacă există probleme pentru care o acţiune paralelă poate ajunge la soluţie mai eficient, în termenii timpului de calcul, decât poate una serială. Cum, în această carte, mă voi ocupa mai ales de probleme de principiu, distincţia dintre calculul paralel şi cel serial nu va avea o semnificaţie prea mare pentru noi.

1.6. Poziţia $\mathcal{C}$ contravine tezei Church-Turing?

Să ne amintim că punctul de vedere $\mathcal{C}$ consideră creierul conştient acţionând într-un mod ce nu poate fi simulat cu nici un fel de algoritm, *top-down* sau *bottom-up* sau de alt tip. Unele persoane, exprimându-şi îndoielile în legătură cu punctul de vedere $\mathcal{C}$, îşi bazează atitudinea pe afirmaţia că $\mathcal{C}$ ar contrazice aşa numita teză Church (sau Church-Turing) care este în general acceptată. Care este teza lui Church? În forma sa originară, formulată de logicianul american Alonzo Church în 1936, teza afirmă că tot ceea ce poate fi considerat în mod rezonabil un proces matematic, 'pur matematic' — adică algoritmic — poate fi realizat pe baza unei scheme particulare descoperită chiar de Church şi numită lambda calcul

[6] *computaţional* — caracteristică a unei activităţi de a putea fi simulată/implementată pe un calculator electronic din zilele noastre (n.t.)

(λ-*calculus*) (o schemă de o deosebită eleganţă şi concizie; vezi MN p. 76-82 pentru o scurtă introducere în subiect). Puţin după aceea, în 1937, matematicianul britanic Alan Turing a găsit propriul său mod de descriere a proceselor algoritmice, în termenii acţiunii unei 'maşini de calcul' teoretice pe care noi o numim azi *maşina Turing*. Cu puţin timp înainte, logicianul american, născut în Polonia, Emil Post (1936) dezvoltase de asemenea o schemă oarecum similară cu cea a lui Turing. Church şi Turing, în mod independent, au arătat echivalenţa dintre calculul lui Church şi maşina Turing. Se consideră că maşinile de calcul actuale s-au dezvoltat într-o măsură foarte mare datorită concepţiilor lui Turing. Aşa cum am menţionat mai sus, o maşină Turing este echivalentă, ca acţiune, cu un calculator modern — cu idealizarea pe care o presupune accesul la o memorie infinită. Astfel, putem privi acum teza lui Church ca afirmaţia care susţine că algoritmii matematici sunt exact ceea ce execută un calculator modern idealizat. Cu siguranţă, nu există nici o contradicţie cu $\mathcal{C}$ atunci când acceptăm această formă a tezei lui Church[7]

S-ar putea ca Turing însuşi să fi avut ceva mai mult în minte şi anume că posibilităţile de calcul ale oricărui dispozitiv *fizic* trebuie să fie (în mod ideal) echivalente cu acţiunea unei maşini Turing. O astfel de afirmaţie, însă, ar depăşi cu mult ceea ce Church pare să fi urmărit. Motivaţia de bază a lui Turing pentru dezvoltarea conceptului de 'maşină Turing' s-a bazat pe ideile sale despre modul în care omul realizează un calcul (vezi Hodges 1983). Se pare că el vedea acţiunea fizică în general — care trebuia să includă şi acţiunea creierului uman — reductibilă la un anumit tip de acţiune a unei maşini Turing. Poate că ar trebui să numim această afirmaţie 'teza lui Turing' pentru a o distinge de 'teza lui Church' care este pur matematică, care nu este în nici un fel contrazisă de $\mathcal{C}$ Aceasta va fi terminologia pe care o voi adopta în acestă carte. În conformitate cu ea, teza lui Turing şi nu cea a lui Church, este cea care vine în contradicţie cu punctul de vedere $\mathcal{C}$.

---

[7] Uneori se constată, în discuţiile matematice, faptul că o procedură este considerată 'evident' algoritmică, prin natura sa, chiar dacă, pe moment, nu se vede cum ar putea fi formulată în termenii unei maşini Turing sau cu ajutorul lambda calculului. În astfel de cazuri, se presupune că o astfel de operaţie trebuie să existe, în conformitate cu "teza lui Church" Vezi de exemplu Cutland (1980). Procedând astfel nu facem nimic nepotrivit şi cu siguranţă nu apare nici o contradicţie cu $\mathcal{C}$. Modul acesta de utilizare a tezei lui Church intră foarte mult în discuţia din capitolul 3.

1.7. Haos

În ultimii ani a stârnit un mare interes fenomenul matematic cunoscut sub numele de *haos*, în care sistemele fizice par să fie capabile de o comportare complet impredictibilă (figura 1.1). Ar putea oare fenomenul numit haos să ne furnizeze baza fizică, necalculabilă, necesară pentru un punct de vedere de natura lui $\mathcal{C}$?

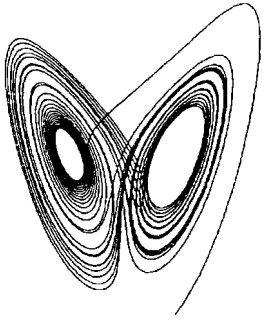


Fig. 1.1. Atractorul Lorentz - vechi exemplu de sistem haotic. Urmând liniile se trece, aparent fără nici o regulă, din lobul stâng în cel drept şi înapoi. Curbă definită de o ecuaţie diferenţială.

Sistemele haotice sunt sisteme fizice evoluând dinamic, simulări matematice ale unor astfel de sisteme fizice, sau, pur şi simplu, modele matematice studiate pentru frumuseţea lor, în care comportarea viitoare a sistemului depinde, într-un mod foarte critic, de condiţiile iniţiale precise ale sistemului.

Deşi sistemele haotice obişnuite sunt complet deterministe şi calculabile, ele pot totuşi, *în practică*, să se comporte de parcă nu ar fi deloc deterministe. Acest lucru se întâmplă din cauza preciziei cu care trebuie cunoscută starea iniţială pentru a putea face o previziune asupra viitorului sistemului, precizie care poate să întreacă tot ceea ce poate fi conceput drept măsurabil.

Un exemplu, citat adeseori, este prognoza detaliată a vremii pe termen lung (prognoza meteo). Legile care guvernează mişcarea moleculelor de aer şi alte cantităţi fizice care ar putea fi relevante pentru calcul sunt perfect cunoscute. Totuşi, modificarea vremii, care poate să aibă loc chiar într-un interval de câteva zile, depinde foarte subtil de condiţiile iniţiale precise, iar aceste condiţii nu pot fi

măsurate suficient de exact, pentru o prognoză sigură. Evident, numărul de parametri care intră într-un astfel de calcul va fi foarte mare şi deci, nu e de mirare că prognoza, în acest caz, poate să se dovedească practic imposibilă.

Pe de altă parte, o comportare haotică poate să apară în sisteme foarte simple, ca acela care constă dintr-un număr mic de particule. Să presupunem, de exemplu că ţi s-ar cere să introduci în gaură o a cincea bilă de biliard, **E**, dintr-un aranjament de bile **A, B, C, D**, plasate pe o linie curbă, prin lovirea cu tacul a bilei **A**, astfel încât **A** să lovească bila **B**, care apoi să lovească bila **C**, apoi **C** să lovească bila **D** şi, în sfârşit, aceasta să lovească bila **E**, care să cadă în gaură. Precizia cerută de o astfel de lovitură este mult mai mare decât abilitatea oricărui jucător de biliard[8]. Dacă în joc sunt 20 de bile, atunci chiar dacă bilele ar fi sfere perfect elastice, introducerea ultimei bile în gaură nu poate fi realizată nici de cea mai precisă maşină produsă cu tehnologia cea mai modernă. Comportarea ultimei bile din şir va fi aleatoare, în ciuda faptului că legile newtoniene, care guvernează mişcarea bilelor sunt complet deterministe şi în principiu, calculabile. Totuşi, nu există nici un calcul care ar putea să prevadă comportarea *de fapt* a ultimei bile din şir, pur şi simplu pentru că nu va exista nici un mijloc de a determina suficient de exact poziţia iniţială şi viteza tacului, sau poziţia bilelor de la începutul şirului. Mai mult, perturbaţii foarte fine, ca de exemplu răsuflarea cuiva din apropiere, pot influenţa precizia atât de mult, încât să facă inutile orice calcule.

Vreau totuşi să fie clar că, în ciuda acestor dificultăţi profunde pentru prezicerea deterministă, toate sistemele normale denumite 'haotice' trebuie incluse în ceea ce eu denumesc 'calculabil'. De ce ? Ca şi în alte situaţii de care ne vom ocupa mai târziu, tot ce avem de făcut pentru a decide dacă o procedură este sau nu calculabilă, este să răspundem la întrebarea: poate să fie pusă pe un calculator? Evident, răspunsul trebuie să fie 'da' în acest caz, pur şi simplu pentru că sistemele haotice descrise matematic sunt studiate, în mod normal, punându-le pe calculator!

Dacă încercăm să realizăm pe calculator o simulare a prognozei detaliate a timpului probabil din Europa pentru o săptămână, sau coliziunile succesive a 20 de bile nealiniate, pe masa de biliard, după o lovitură rapidă de tac, atunci

---

[8] Într-o formă mai veche a acestui text nu precizasem că bilele să fie aşezate pe o curbă. Făcând nişte încercări, pentru bilele aşezate pe o dreaptă, lovitura se dovedeşte uşoară. Dacă iniţial bilele sunt aliniate apare o stabilitate care nu are loc în cazul general.

simularea noastră, aproape sigur, nu va evolua deloc aşa cum se petrec lucrurile în *realitate*. Acest lucru este în natura sistemelor haotice. Nu se poate face o prezicere pe bază de calcul a comportării *actuale* a sistemului. Fără îndoială, simularea unei comportări *tipice* a sistemului este perfect posibilă.

Timpul probabil, prognozat poate foarte bine să nu semene cu vremea de afară, deşi este o descriere perfect plauzibilă a vremii. La fel, prezicerea comportării bilelor în coliziune pe masa de biliard poate fi perfect acceptabilă, ca variantă de comportare posibilă, chiar dacă nu are nimic de a face cu situaţia reală de pe masa de biliard. Un alt aspect care subliniază natura perfect calculabilă a acestor operaţii este faptul că reluarea unui anumit calcul cu aceleaşi date de intrare pentru simulare duce la aceleaşi rezultate ca mai înainte (acest lucru presupune ca nici calculatorul să nu facă vreo eroare; dar calculatoarele moderne fac foarte rar greşeli).

În contextul inteligenţei artificiale nu se încearcă simularea comportării tuturor individualităţilor particulare; ar fi complet satisfăcătoare chiar şi numai simularea comportării unei singure individualităţi! Astfel, nu e lipsit de raţiune să adoptaţi punctul meu de vedere: sistemele haotice trebuie cu siguranţă incluse în ceea ce am numit *computaţional*. Simularea pe calculator a unui astfel de sistem va furniza un 'caz tipic' perfect raţional, chiar dacă se va constata că nu este şi un 'caz actual' Dacă manifestările exterioare ale inteligenţei umane sunt rezultatul unei anumite evoluţii dinamice haotice — o evoluţie calculabilă în sensul descris deja — atunci acest lucru va fi în concordanţă cu punctele de vedere $\mathcal{A}$ şi $\mathcal{B}$, dar nu şi cu $\mathcal{C}$.

Uneori s-a sugerat că acest fenomen de haos, dacă intervine în activitatea internă a creierului fizic, ar fi ceea ce ar permite creierului nostru să se comporte într-un mod care să *apară* ca diferit de activitatea deterministă a unei maşini Turing, chiar dacă, aşa cum am arătat mai sus, este, tehnic vorbind, calculabil. Va trebui să revin asupra acestui subiect în paragraful 3.22.

Pentru moment, ceea ce trebuie să fie clar este faptul că sistemele haotice *sunt* incluse în ceea ce am înţeles prin *calculabil* sau *algoritmic*. Întrebarea, dacă ceva poate fi simulat *practic*, este diferită de problema *de principiu* pe care am considerat-o aici.

1.8. Calcul analogic

Până aici am considerat *calculul* numai în sensul în care acest termen se aplică calculatoarelor digitale moderne sau, mai precis, precursorului lor teoretic: maşina Turing. Dar mai există şi alte tipuri de dispozitive de calcul care s-au folosit mai demult, în care operaţiille nu sunt realizate prin stări discrete *on/off* familiare pentru calculatoarele *digitale*, ci cu ajutorul unor parametri fizici continui. Cel mai familiar dispozitiv de acest fel este rigla de calcul, unde parametrul fizic este distanţa liniară (de-a-lungul riglei). Această distanţă este folosită pentru a reprezenta logaritmii numerelor care urmează să fie înmulţite sau împărţite. Există multe tipuri diferite de dispozitive de calcul analogice, utilizând alţi parametri fizici, ca timpul, masa, sau potenţialul electric.

Sistemele analogice ne pun anumite probleme tehnice deoarece noţiunea standard de calcul şi calculabilitate se aplică numai sistemelor *discrete* (cele care lucrează cu operaţii 'digitale') şi nu continue, cum sunt distanţa sau potenţialul electric, să zicem, care sunt proprii fizicii clasice convenţionale. Pentru a aplica noţiunea obişnuită de calculabilitate unui sistem a cărui descriere cere parametri continui şi nu discreţi, este natural să apelăm la aproximări. Într-adevăr, la simularea cu calculatorul a sistemelor fizice, în general, procedura obişnuită este să se aproximeze toţi parametrii continui prin discretizare. Apar în acest caz anumite erori, iar dacă se alege o anumită precizie pentru aproximare, ar putea exista sisteme fizice pentru care această aproximare să nu fie suficientă. Astfel, simularea pe calculator ar putea duce la concluzii eronate în privinţa comportării sistemului fizic care a fost simulat.

În principiu, precizia poate fi totdeauna mărită până când ajunge să fie adecvată sistemului continuu ce trebuie simulat. Totuşi, mai ales când este vorba despre sisteme haotice, timpul total de calcul şi memoria cerută pot deveni prohibitive în practică. În plus, există un aspect tehnic şi anume, nu putem fi niciodată siguri că precizia aleasă *este* suficientă. Vor fi necesare anumite teste care să semnaleze când o rafinare a preciziei nu mai este necesară şi când comportarea calitativă calculată cu acest nivel de precizie este adecvată. Aceste lucruri ridică anumite probleme delicate de matematică şi nu voi intra acum în detalii.

Există şi alte abordări ale problemelor ridicate de calculabilitatea sistemelor discrete, în care sistemele sunt tratate ca structuri matematice, având propria lor

noţiune de calculabilitate — o noţiune care generalizează ideea de calculabilitate Turing de la discret la continuu [12]. Utilizând o astfel de noţiune, nu mai este necesară aproximarea sistemelor continue prin discretizarea parametrilor pentru a putea fi aplicată noţiunea convenţională de calculabilitate Turing. Aceste idei sunt interesante din punct de vedere matematic, dar din nefericire ele nu par să fi atins încă naturaleţea şi unicitatea care se aplică noţiunii standard de calculabilitate Turing pentru sisteme discrete. Mai mult, apar anumite anomalii, printre care o 'necalculabilitate' pentru sisteme simple, unde nu este clar dacă o astfel de terminologie este potrivită (de exemplu, chiar pentru simpla ecuaţie a undei, din fizică; vezi Pourel şi Richards (1981), MN p. 205). Pe de altă parte, trebuie menţionată lucrarea lui Rubel(1989) unde, pentru calculatoarele analogice dintr-o anumită clasă, destul de restrânsă, se arată că nu pot ajunge dincolo de calculabiltatea Turing obişnuită. Cred că acestea sunt probleme interesante care vor fi clarificate de cercetări viitoare. Oricum, pentru mine nu e clar dacă aceste cercetări au atins sau nu punctul de la care să poată fi aplicate problemelor puse în discuţie aici.

În această carte, mă preocup, în mod special, de problema naturii calculabile a acţivităţii mentale, unde termenul *calculabilitate* este luat în sensul obişnuit de *calculabilitate Turing*.. Calculatoarele obişnuite din zilele noastre sunt prin natura lor, digitale — numerice şi, pentru activitatea IA de azi, acest lucru este relevant. Se prea poate ca, în viitor, să apară calculatoare de un tip diferit, care să facă critică utilizarea parametrilor fizici şi care să le permită să se comporte complet *diferit* de calculatoarele digitale.

Toate acestea au importanţă pentru distincţia dintre versiunea *slabă* şi cea *tare* a punctului de vedere $\mathcal{C}$. Conform versiunii *slabe* a lui $\mathcal{C}$, trebuie să existe acţiuni fizice, ce stau la baza comportării creierului uman conştient şi care nu sunt calculabile în sensul standard al calculabilităţii Turing, dar care pot fi pe de-a-ntregul înţelese în termenii teoriilor din fizica zilelor noastre. Pentru ca aşa ceva să fie posibil, trebuie ca aceste acţiuni să depindă de parametri fizici continui, astfel încât aceştia să nu poată fi simulaţi prin metode standard digitale. Pe de altă parte, în conformitate cu versiunea *tare* a lui $\mathcal{C}$, necalculabilitatea s-ar datora unei teorii fizice non-computaţionale — încă nedescoperită — ale cărei implicaţii ar fi esenţiale pentru acţiunea creierului conştient. Deşi această a doua posibilitate ar putea să pară impresionantă, alternativa (pentru suporterii $\mathcal{C}$) este să se

găsească rolul unora dintre acţiunile continui din legile cunoscute ale fizicii, care nu pot fi simulate prin calcul în mod corespunzător. Totuşi, pentru moment, ceea ce se poate spera pentru un sistem analogic de orice tip, care a fost serios studiat până acum, ar fi să i se furnizeze o simulare digitală efectivă.

În afara acestor consideraţii teoretice generale, calculatoarele digitale de astăzi s-au dovedit mai avantajoase decât cele analogice. Acţiunea digitală este mult mai precisă, în principal pentru că o memorie digitală permite creşterea preciziei simplu, prin mărirea lungimii alocate reprezentării numerelor, lucru ce se obţine uşor prin creşterea modestă (logaritmică) a capacităţii calculatorului; în vreme ce pentru maşinile analogice (cel puţin pentru cele complet analogice, în care nu sunt importate concepte digitale), creşterea preciziei de calcul se realizează printr-o creştere (liniară) enormă a capacităţii calculatorului. În viitor s-ar putea să apară idei noi care să fie în avantajul maşinilor analogice dar, cu tehnologia de azi, cele mai multe avantaje practice semnificative sunt de partea calculatoarelor numerice.

1.9. Ce fel de acţiune ar putea fi non-computaţională?

Cele mai multe acţiuni care ne vin în minte sunt lucruri care ar trebui incluse în cele pe care le-am numit *computaţionale* (în sensul de *calculabile cu un calculator*). Cititorul s-ar putea teme că nu rămâne nimic rezonabil cu care să opereze afirmaţia $\mathcal{C}$. Dar nu am spus nimic până acum despre acţiunile *aleatoare*, care ar putea fi furnizate, să zicem, de un sistem cuantic (mecanica cuantică va fi discutată pe larg în partea a doua, în capitolele 5 şi 6). Totuşi, este greu de văzut care ar fi avantajele unui sistem care ar avea la intrare date aleatoare, faţă de unul care ar lucra cu date *pseudo-aleatoare*, generate de un calculator (vezi paragraful 3.11). Deşi există anumite diferenţe tehnice între *aleator* şi *pseudo-aleator*, aceste diferenţe par să nu aibă importanţă pentru **IA**. În paragrafele 3.11, 3.18 şi următoarele, voi aduce argumente tari pentru a arăta că 'aleatorul pur' nu ne oferă nimic util din punctul nostru de vedere; şi atunci e mai bine să rămânem la comportarea haotică pseudo-aleatoare, care trece drept calculabilă.

Ce se poate spune despre rolul mediului? Pe măsură ce o fiinţă umană se dezvoltă, ea are un mediu înconjurător unic, neîmpărţit cu nimeni altcineva. Oare

nu s-ar putea spune că acest mediu personal, unic, este cel care ne dă date de intrare necalculabile? Îmi este totuşi greu să văd în ce fel 'unicitatea' mediului nostru ne-ar putea ajuta în acest context. Discuţia este asemănătoare cu cea despre haos (ca în paragraful 1.7). Presupunând că nu există nimic mai mult decât calcul în simularea unui mediu haotic *plauzibil*, o astfel de simulare este tot ce ne trebuie pentru a antrena un robot controlat de un calculator. Robotul nu are nevoie să înveţe nici un truc pentru a evolua într-un anumit mediu; un mediu *tipic* simulat (mai degrabă decât unul real) este cu siguranţă suficient pentru robot.

Există ceva, într-un mediu exterior, care să fie în mod inerent imposibil de simulat cu calculatorul? Poate că *există* în lumea fizică exterioară ceva situat în afara posibilităţilor de simulare prin calcul. Unii susţinători ai punctelor de vedere $\mathcal{A}$ şi $\mathcal{B}$ ar putea fi înclinaţi să atribuie unei deficienţe de calculabiltate a mediului înconjurător aşa-zisa necalculabilitate existentă în comportamentul uman. Ar fi imprudent pentru aceşti suporteri să se bazeze pe astfel de argumente. Momentan se acceptă că ar putea să existe ceva, *undeva*, în comportarea fizică, care să nu poată fi simulat prin calcul, iar acest lucru ar putea zdruncina motivul principal de îndoială în plauzibilitatea lui $\mathcal{C}$. Dacă există acţiuni în mediul exterior care nu pot fi simulate cu calculatorul, atunci de ce n-ar exista unele şi în interiorul creierului? Organizarea fizică internă a creierului uman este, în definitiv, mai complicată decât cele mai multe medii, cu excepţia, probabil, a unor medii puternic influenţate de acţiunile altor creiere umane. Acceptarea existenţei unor acţiuni fizice externe necalculabile exclude principalul argument împotriva lui $\mathcal{C}$ (vezi de asemenea discuţia din paragrafele 3.9 şi 3.10).

Mai trebuie să fac o observaţie în legătură cu ceea ce înţeleg prin 'a depăşi nivelul computaţional' aşa cum cere $\mathcal{C}$ Mă gândesc nu numai la ceva care devansează calculul *practic* Orice simulare a unui mediu plauzibil sau a proceselor fizice şi chimice din creier ar putea să pară ceva care, în principiu, este calculabil, doar că durează prea mult şi necesită un spaţiu de memorie prea mare, astfel încât nu există speranţa ca un astfel de calcul să fie dus la bun sfârşit vreodată de vreun calculator actual sau viitor. Evident că scrierea unui astfel de program de calcul este de neimaginat datorită numărului mare de factori de care trebuie să ţinem seama. Oricât de importante ar fi aceste consideraţii (şi le vom discuta atent în răspunsul la Q8 din paragraful 2.6 precum şi în paragraful 3.5), ele nu se referă la ceea ce înţeleg prin non-computabilitate, în sensul lui $\mathcal{C}$

Dimpotrivă, ceea ce înţeleg este ceva aflat, *în principiu*, dincolo de calcul, într-un sens pe care îl voi descrie îndată. Calculele aflate pur şi simplu dincolo de posibilităţile de calcul ale calculatoarelor (actuale sau viitoare) sunt, încă, în sens teoretic, calcule.

Cititorul s-ar putea întreba, pe bună dreptate: dacă nu există nimic care să conteze ca 'non-computaţional', nici în evenimentele întâmplătoare, nici în influenţele mediului, nici în complicaţiile de negestionat, atunci ce anume aş putea eu să-mi imaginez ca non-computaţional — aşa cum cere punctul de vedere $\mathcal{C}$? Ceea ce am eu în vedere se referă la o activitate matematică precisă despre care să se poată demonstra că este mai presus de calcul. După câte ştim deocamdată, nu este nevoie de o astfel de activitate pentru a descrie comportarea fizică. Cu toate acestea, o posibilitate logică există; chiar mai mult decât o posibilitate logică. În conformitate cu argumentele cărţii de faţă, ceva necalculabil trebuie să fie inerent legilor fizicii, în ciuda faptului că astfel de lucruri nu se întâlnesc încă în fizica actuală. Anumite exemple de activitate matematică, de tipul celor la care mă gândesc, sunt deosebit de simple, aşa încât voi folosi unele dintre ele ca ilustraţii.

Trebuie să încep prin a descrie câteva exemple de clase de probleme matematice bine definite, care — într-un sens pe care îl voi preciza îndată — nu au o soluţie generală calculabilă. Pornind cu oricare dintre aceste clase de probleme, vom putea să construim 'un model jucărie' al unui Univers fizic a cărui acţiune, deşi complet deterministă, nu poate fi simulată prin calcul.

Primul dintre exemple este şi cel mai faimos este cunoscut ca 'problema a zecea a lui Hilbert', care a fost enunţată de marele matematician David Hilbert în 1900, una dintre problemele dintr-o listă de întrebări, atunci încă fără răspuns şi care au însemnat foarte mult pentru dezvoltarea ulterioară a matematicii din secolul douăzeci. Problema a zecea a lui Hilbert cerea să se găsească un procedeu de calcul pentru a decide, pentru un sistem de ecuaţii diofantiene, dacă ecuaţiile au vreo soluţie comună.

Ce sunt ecuaţiile diofantiene ? Sunt ecuaţii polinomiale, cu oricâte variabile, având coeficienţii şi toate soluţiile, numere întregi. (Un număr întreg este unul dintre numerele din lista ..., -3, -2, -1, 0, 1, 2, 3, 4, Ecuaţiile diofantiene au fost pentru prima oară studiate sistematic de matematicianul grec Diophantos în secolul III e.n.) Un exemplu de sistem de ecuaţii diofantiene este

$$6w+2x^2-y^3=0, \qquad 5xy-z^2+6=0, \qquad w^2-w+2x-y+z-4=0$$

Un alt exemplu este:

$$6w+2x^2-y^3=0, \qquad 5xy-z^2+6=0, \qquad w^2-w+2x-y+z-3=0.$$

Primul sistem are ca soluţie, în particular, pe

$$w=1, \quad x=1, \quad y=2, \quad z=4,$$

în vreme ce al doilea sistem nu are soluţie deloc (deoarece, din prima dintre ecuaţii, y trebuie să fie un număr par, iar din a doua, z trebuie să fie par de asemenea, ceea ce contrazice ecuaţia a treia, oricare ar fi w, deoarece w^2-w este totdeauna un număr par şi 3 este un număr impar). Problema pusă de Hilbert era să se găsească o procedură matematică — un *algoritm* — pentru a decide care sisteme diofantiene au soluţii — aşa cum a avut primul sistem de mai sus — şi care nu, ca în al doilea exemplu. Să ne amintim că un algoritm este un procedeu de calcul — acţiunea unei maşini Turing (vezi paragraful 1.5). Deci, problema a zecea a lui Hilbert cere o procedură de calcul pentru a decide când un sistem de ecuaţii diofantiene poate fi rezolvat.

A zecea problemă a lui Hilbert a fost, istoric vorbind, foarte importantă, deoarece a ridicat la rândul ei chestiuni noi, care nu fuseseră niciodată puse mai înainte. Ce înseamnă, de fapt, în termeni matematici precişi, că o problemă are o soluţie algoritmică ? Ce este, de fapt, un algoritm ? Aceasta a fost întrebarea care l-a condus pe Alan Turing, în 1936, să propună propria lui definiţie pentru ceea ce înseamnă, în terminologia actuală, maşina Turing. Alţi matematicieni (Church, Kleene, Godel, Post şi alţii; vezi Gandy (1988)) au propus, cam în acelaşi timp, proceduri diferite, dar ulterior s-a constatat că abordarea lui Turing a fost cea mai influentă. (El a fost singurul care a avut ideea de a introduce o maşină algoritmică particulară 'care să includă totul' — pe care a numit-o maşină Turing *universală* — care poate, ea însăşi, să îndeplinească orice acţiune algoritmică. Această idee a condus la calculatorul cu destinaţie universală care ne este acum atât de familiar.) Turing a putut arăta că există anumite clase de probleme care *nu au* soluţie algoritmică (în particular *problema opririi*, pe care o voi descrie mai jos). Totuşi, problema a zecea a lui Hilbert a trebuit să aştepte până în 1970 când matematicianul rus Yuri Matiyasevich — furnizând o demonstraţie care completa unele argumente susţinute mai înainte de americanii Julia Robinson, Martin Davis şi Hilary Putnam — a arătat că nu poate exista vreun program de calculator (algoritm) care să decidă, sistematic, cu da/nu dacă un sistem de ecuaţii diofantiene

are o soluţie (vezi Davis (1978) şi Devlin (1988), capitolul 6, pentru o prezentare a acestei poveşti). Trebuie remarcat faptul că ori de câte ori răspunsul este 'da', la el se poate ajunge, în principiu, cu un program de calculator care încearcă muncitoreşte toate numerele întregi, unele după altele. Pe de altă parte, răspunsul 'nu' rezistă oricărei tratări sistematice. Pot fi date o mulţime de reguli pentru a obţine corect răspunsul 'nu' — ca argumentul cu numerele pare şi impare pe care l-am utilizat pentru a soluţiona al doilea sistem dat mai înainte — dar, teorema lui Matiyasevich arată că acest procedeu nu poate fi *niciodată* complet.

Un alt exemplu, de problemă bine definită matematic, care nu are o soluţie algoritmică, este *problema acoperirii*. Ea se formulează astfel: fiind dată o mulţime de forme poligonale, să se decidă dacă cu aceste forme se poate acoperi planul; cu alte cuvinte, este posibil să acoperim întreg planul euclidian utilizând doar aceste forme, fără goluri şi fără suprapuneri? Matematicianul american Robert Berger, în 1966, a arătat că această problemă este computaţional insolubilă, utilizând o extindere a unei lucrări anterioare, din 1961, a matematicianului chinezo-american Hao Wang (vezi Grunbaum şi Shephard 1987). De fapt, aşa cum am formulat problema, sunt unele imprecizii, deoarece formele poligonale trebuie să fie precizate într-un anumit fel utilizând numerele reale (numerele care se definesc cu un număr infinit de zecimale), în vreme ce algoritmii obişnuiţi operează cu numere întregi[9] Această imprecizie poate fi evitată presupunând elemente de acoperire constând dintr-un număr de pătrate alăturate prin muchii. Astfel de elemente de acoperire se numesc *polyominoes* (vezi Golomb 1965; Gardner 1965, capitolul 13; Klarner 1981). Dăm câteva exemple în figura 1.2. (pentru alte exemple de mulţimi de acoperire vezi **MN** p. 148-154 şi fig. 4.6-4.12.). Este curios faptul că insolvabilitatea computaţională a problemelor de acoperire depinde de existenţa unei mulţimi de acoperire numită mulţime aperiodică — care acoperă planul doar non-periodic (adică, într-un mod în care partea acoperită a planului nu se repetă, oricât de departe s-ar extinde acoperirea). În figura 1.3 este prezentată o mulţime de trei elemente de acoperire (dezvoltate dintr-o acoperire descoperită de Robert Ammann în 1977 conform cu Grunbaum şi Shephard (1987), figurile 10.4.11-10.4.13 de la p. 555-556).

---

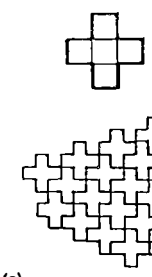
[9] De fapt,algoritmii şi cu atât mai mult implementările lor pe calculator, operează cu aproximări cu un număr finit de cifre (binare, în calculator) ale numerelor reale (n.t.)

Faptul că problema a zecea a lui Hilbert sau problema acoperirii nu pot fi soluţionate prin calcul are o demonstraţie matematică dificilă şi nu voi încerca aici să o expun [13]. Acest tip de problemă se reduce la una pe care Turing a abordat-o în lucrarea sa originală: insolvabilitatea prin calcul a *problemei opririi* — problema de a decide situaţiile în care acţiunea unei maşini Turing nu ajunge să se oprească. În paragraful 2.3, vor fi date exemple variate în care calculele nu se opresc niciodată; şi în paragraful 2.5 va fi prezentată o argumentaţie relativ simplă — bazată pe demonstraţia originală a lui Turing — care arată , printre alte lucruri, că problema opririi este insolubilă prin calcul (implicaţiile acelor 'alte lucruri' pe care argumentaţia le conţine vor fi centrale în tot ce se discută în partea întâi a cărţii!).

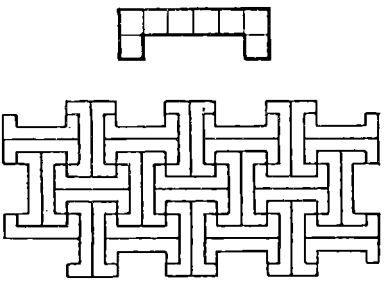
Cum poate fi utilizată o astfel de clasă de probleme ca ecuaţiile diofantiene sau problema acoperirii, pentru a construi un Univers jucărie care să fie determinist şi în acelaşi timp necalculabil?

Să presupunem că modelul nostru de Univers are un *timp discret*, parametrizat prin numere naturale (întregi nenegativi) 0, 1, 2, 3, 4, La momentul n, starea Universului este specificată printr-o clasă de probleme considerată — să zicem, printr-o mulţime de elemente de acoperire. Ar trebui să existe două reguli bine definite care să stabilească care dintre mulţimile de *polyomino* reprezintă starea Universului la momentul $n+1$, cunoscând mulţimea de *polyomino* care reprezintă Universul la starea n. Prima dintre aceste reguli ar trebui adoptată dacă elementele *acoperă* planul şi a doua regulă să fie adoptată dacă ele *nu acoperă* planul. Detaliile despre cum pot fi formulate aceste reguli nu sunt importante. S-ar putea forma o listă S_0, S_1, S_2, S_3,... cu toate mulţimile de *polyomino* astfel încât cele care conţin un număr par de pătrate să aibă indice par: S_0, S_2, S_4,...; iar cele care conţin un număr impar de pătrate să aibă indice impar: S_1, S_3, S_5,.... (un astfel de aranjament nu este prea greu de făcut utilizând o procedură de calcul). 'Evoluţia dinamică' a modelului nostru de Univers jucărie este dată acum de:

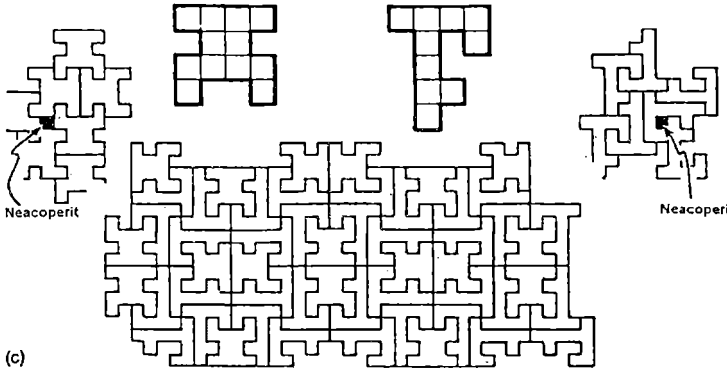
> Starea S_n a Universului la momentul t trece în starea S_{n+1} la momentul $t+1$, dacă mulţimea S_n *acoperă* planul şi S_{n+2} dacă mulţimea S_n *nu acoperă* planul.



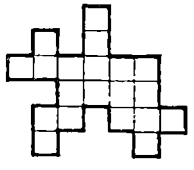
(a)

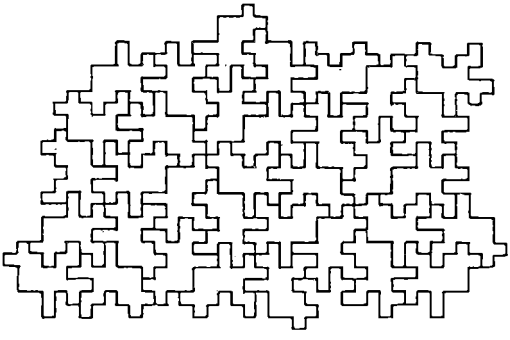


(b)



(c)





(d)

Fig. 1.2. Câteva mulţimi de *polyominoes* care acoperă planul euclidian (reflectarea fiind admisă). Dar nici una din formele din setul (c) nu poate acoperi singură planul!

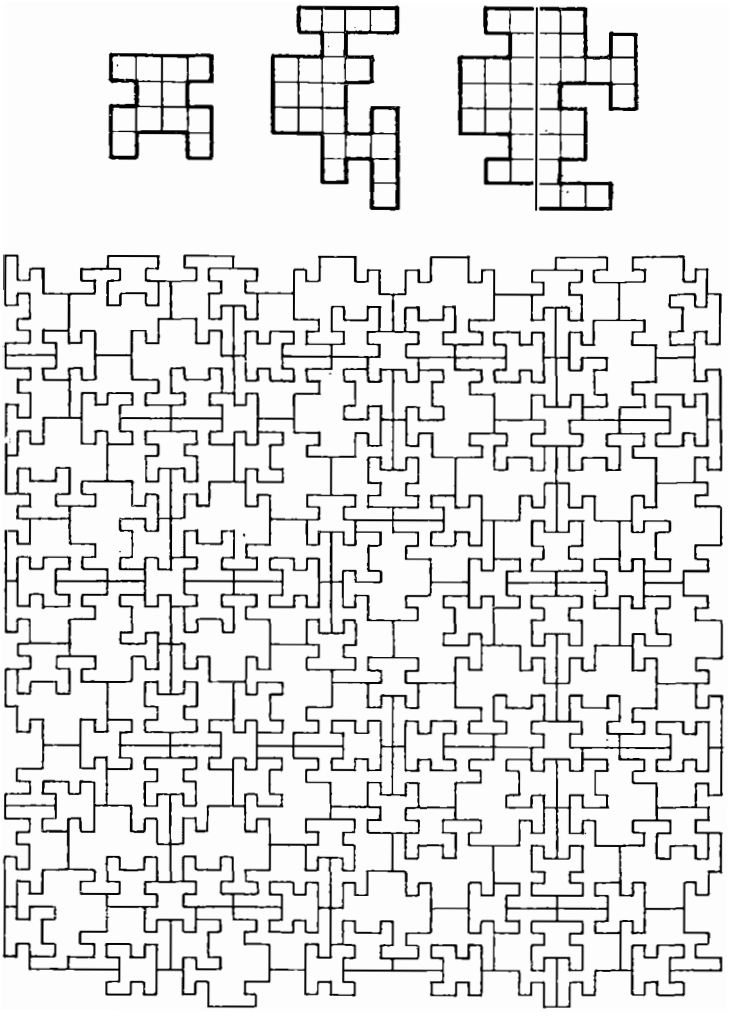


Fig. 1.3. O mulţime de trei 'polyominoes' care acoperă planul, dar numai într-un mod total nerepetitiv (derivat dintr-o mulţime a lui Robert Ammann)

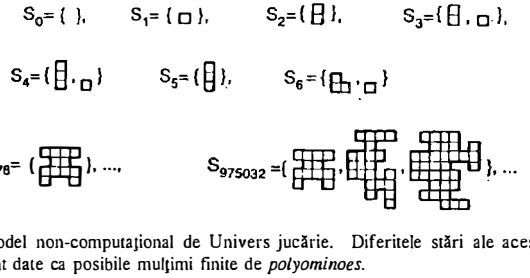


Fig. 1.4. Un model non-computaţional de Univers jucărie. Diferitele stări ale acestui Univers deterministic sunt date ca posibile mulţimi finite de *polyominoes*.

Un astfel de *Univers* se comportă complet determinist, dar, cum nu există o procedură de calcul care să stabilească când o mulţime S_n acoperă planul (sau echivalent, când numărul total de pătrate este par, sau impar), nu există o simulare prin calcul a dezvoltării sale. (Vezi figura 1.4. Elementele de acoperire, în acest caz, sunt numerotate astfel încât numerele pare S_n corespund la un număr total de pătrate par, iar numerele impare la număr impar de pătrate. Evoluţia în timp are loc în ordinea numerică S_0, S_2, S_3, S_4,..., S_{278}, S_{280},..., doar că un număr este sărit oridecâte ori mulţimea precedentă nu acoperă planul.)

Evident, o astfel de schemă nu poate fi luată în serios când este vorba de modelarea Universului pe care îl locuim. Am folosit-o aici pentru a ilustra faptul prea puţin înţeles că există o diferenţă clară între determinism şi calculabilitate.

Există modele complet deterministe de Univers, cu reguli de evoluţie clar definite, care sunt imposibil de simulat prin calcul.

De fapt, aşa cum vom vedea în paragraful 7.9, modelele de un tip foarte particular, pe care le-am considerat aici, se vor dovedi insuficiente pentru ceea ce ne trebuie în justificarea punctului de vedere $\mathcal{C}$ Dar se va vedea în paragraful 7.10 că există totuşi unele posibilităţi fizice interesante pentru ceea ce *avem* nevoie!

1.10. Ce se poate spune despre viitor?

Ce ne indică, pentru viitorul acestei planete punctele de vedere $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$? În conformitate cu $\mathcal{A}$ va urma un moment în care supercalculatoare programate corespunzător vor atinge — şi apoi vor depăşi — toate capacităţile mentale ale oamenilor. Evident, diferitele persoane care susţin punctul de vedere

$\mathcal{A}$ apreciază în moduri diferite momentul când acest lucru se va întâmpla. Unii sunt de părere că vor trece multe secole până când calculatoarele vor atinge nivelul nostru, deoarece se ştie prea puţin despre calculele pe care creierul trebuie să le facă (cred ei) pentru a atinge subtilitatea acţiunilor noastre, — o subtilitate care va fi necesară înainte ca o 'conştienţă' apreciabilă să fie atinsă. Alţii vorbesc despre un interval de timp mult mai scurt. Hans Moravec, în cartea sa, *Mind Children* (1988), construieşte un scenariu — bazat pe ritmul de dezvoltare al tehnologiei calculatoarelor în ultima jumătate de secol şi pe proporţia din activitatea creierului despre care consideră că a fost deja simulată cu succes — pentru a susţine că 'echivalenţa umană' va fi atinsă cam prin anul 2030 (sunt şi alţii care susţin termene şi mai apropiate [14] — câteva din acestea au şi trecut!).

Ca nu cumva cititorul să se simtă cuprins de descurajare în faţa perspectivei de a fi întrecut de calculatoare în mai puţin de, să zicem, 40 de ani, o speranţă este dată — de fapt promisă — de perspectiva de a transfera 'programele din mintea noastră' în corpurile strălucitoare, de plastic sau de metal, ale roboţilor, astfel obţinând pentru noi înşine o formă de nemurire (Moravec 1988,1994).

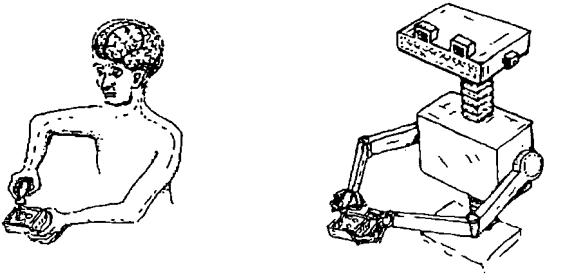


Fig. 1.5. Conform cu $\mathcal{B}$ simularea activităţii umane conştiente fiind posibilă, roboţii vor putea depăşi performanţele omului.

Un astfel de optimism nu stă la îndemâna susţinătorilor punctului de vedere $\mathcal{B}$. Poziţia lor nu diferă de $\mathcal{A}$ cu privire la ceea ce pot calculatoarele, în cele din urmă, să facă privite din exterior. O simulare adecvată a acţiunii creierului uman poate (ea însăşi) să fie utilizată pentru a controla un robot, simularea fiind tot ceea ce este necesar (figura 1.5).

Problema prezenţei conştiinţei conştiente, care se ridică în legătură cu

simularea, este, în conformitate cu $\mathcal{B}$, irelevantă pentru modul în care se comportă roboţii. S-ar putea să dureze secole, sau mai puţin de 40 de ani, ca o astfel de simulare să fie, tehnic, posibilă. Iar, în conformitate cu $\mathcal{B}$, ea ar trebui să fie posibilă. Atunci, acele calculatoare vor atinge nivelul 'echivalenţei umane'; şi s-ar putea chiar să-l depăşească. Opţiunea de a ne 'alătura' roboţilor comandaţi de calculator nu ne mai este deschisă şi se pare că va trebui să ne resemnăm cu perspectiva unei planete guvernate, în final, de maşini insensibile! Dintre toate variantele $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$, mie mi se pare că varianta $\mathcal{B}$ oferă cea mai pesimistă perspectivă a viitorului planetei noastre — în ciuda aparentului său 'bun simţ'!

Pe de altă parte, în conformitate cu $\mathcal{C}$ şi $\mathcal{D}$ ne putem aştepta ca, în viitor, calculatoarele să rămână mereu sub controlul nostru, oricât de departe ar avansa ele în privinţa vitezei, capacităţii şi proiectării logice.

Punctul de vedere $\mathcal{C}$ este deschis în privinţa dezvoltării viitoare a ştiinţei, care ar putea conduce la dispozitive — care să nu se bazeze pe calculatoare, aşa cum le înţelegem noi astăzi, ci pe acţiuni fizice non-computaţionale, despre care $\mathcal{C}$ susţine că stau la baza proceselor noastre de gândire conştientă — dispozitive care să poată atinge cu *adevărat* inteligenţa şi conştienţa.

Poate că aceste dispozitive, nu calculatoarele, aşa cum le înţelegem noi astăzi, vor fi acelea care vor depăşi, eventual, capacităţile omului. Este foarte posibil, dar astfel de speculaţii par premature în acest moment, deoarece ne lipseşte aproape complet înţelegerea ştiinţifică necesară, fără a mai vorbi despre elementele tehnologice. Voi reveni la aceasta în partea a doua a cărţii (paragraful 8.1).

1.11. Pot calculatoarele să aibă drepturi şi responsabilităţi ?

O problemă legată de cele spuse mai sus — şi care ar putea avea o importanţă practică imediată — a început să atragă atenţia teoreticienilor din domeniul *dreptului* [15]. Întrebarea este dacă s-ar putea susţine că, într-un viitor nu prea îndepărtat, calculatoarele ar trebui să aibă responsabilităţi legale, sau anumite drepturi. Cu siguranţă, dacă, într-un timp previzibil, calculatoarele se vor apropia, sau chiar vor depăşi nivelul de expertiză uman, în multe aspecte ale vieţii, atunci întrebările de acest tip vor deveni actuale.

Dacă cineva crede în punctul de vedere $\mathcal{A}$, acela cu siguranţă va accepta că roboţii conduşi de calculatoare (sau calculatoarele) ar trebui să aibă, potenţial, atât drepturi cât şi responsabilităţi, deoarece în concordanţă cu acest punct de vedere, între noi şi roboţii suficient de evoluaţi nu există nici o diferenţă esenţială — în afara unor 'accidente' privind materialul de construcţie. Pentru cei care susţin punctul de vedere $\mathcal{B}$ chestiunea este totuşi mai puţin clară. Se poate argumenta în mod rezonabil că mânia, răzbunarea, cinstea, intenţia, adevărul, credinţa, pasiunea sau înţelegerea — sunt esenţiale când este vorba de drepturi şi responsabilităţi. În conformitate cu $\mathcal{B}$, un robot controlat de un calculator nu ar avea nici una dintre aceste calităţi şi cred că de aceea s-ar putea să nu aibă nici drepturi nici responsabilităţi. Mai mult, în conformitate cu $\mathcal{B}$, nu există nici o modalitate efectivă de a susţine că aceste calităţi lipsesc, aşa încât am putea fi puşi în încurcătură, dacă roboţii ar putea imita suficient de bine comportamentul uman.

Această încurcătură pare să fie eliminată în cazul punctului de vedere $\mathcal{C}$ (şi probabil şi în cazul $\mathcal{D}$) deoarece, în conformitate cu acesta, calculatoarele nu pot etala, în mod convingător, calităţi mentale — şi cu siguranţă nici nu le posedă.

Rezultă că niciodată calculatoarele nu vor avea nici drepturi şi nici responsabilităţi. Pentru mine, acesta este un punct de vedere foarte rezonabil, bun de adoptat. În această carte voi pleda cu tărie împotriva punctelor de vedere $\mathcal{A}$ şi $\mathcal{B}$ Acceptarea argumentelor mele simplifică foarte mult poziţia legală: calculatoarele sau roboţii comandaţi de calculatoare nu trebuie să aibă niciodată drepturi şi responsabilităţi. Mai mult, calculatoarele nu merită ruşinea unui blam atunci când lucrurile merg rău — căci,responsabilităţile se află altundeva!

Dar, să fie clar, aceste argumente nu s-ar aplica unor ipotetice 'dispozitive' care ar beneficia de o fizica non-computaţională. Cum perspectiva unor astfel de dispozitive — dacă vor putea fi vreodată construite — nu se vede încă la orizont, nu există nici o problemă de legalitate pentru viitorul previzibil.

Chestiunea 'responsabilităţii' ridică întrebări filosofice adânci în legătură cu cauza ultimă a comportamentului nostru. S-ar putea argumenta că oricare dintre acţiunile noastre este finalmente determinată de ereditatea noastră şi de mediu — sau altfel spus, de acei numeroşi factori întâmplători care ne influenţează viaţile. Nu sunt *toate* aceste influenţe 'dincolo de controlul' nostru, deci lucruri de care nu putem fi în cele din urmă responsabili?

Este problema 'responsabilităţii' numai una de terminologie convenţională,

sau există altceva acolo — un 'sine' aflat deasupra tuturor acestor influenţe — care exercită controlul asupra actelor noastre ? Problema legală a 'responsabilităţii' *pare* să implice ceva ce este cu adevărat în fiecare dintre noi, un fel de 'sine' independent cu *propriile* sale responsabilităţi — şi prin implicaţie, drepturi — a cărui acţiune nu poate fi atribuită nici eredităţii, nici mediului şi nici întâmplării. Dacă este altceva decât pur şi simplu o convenţie de limbaj când vorbim despre un 'sine' independent, atunci trebuie să existe un ingredient care lipseşte din înţelegerea noastră fizică din momentul de faţă. Descoperirea unui astfel de ingredient ar modifica cu siguranţă, profund, perspectiva noastră ştiinţifică.

Cartea de faţă nu dă un răspuns acestor probleme, dar cred că este în stare să ofere o deschidere. Ea nu ne va spune că trebuie să existe cu necesitate un 'sine' ale cărui acţiuni nu pot fi atribuite unor cauze externe, dar ne va îndemna să lărgim conceptia despre ceea ce înseamnă 'cauză'. O 'cauză' ar putea să fie ceva care să nu poată fi calculat nici în practică şi nici în principiu. Voi arăta că, atunci când o 'cauză' apare în urma acţiunilor noastre conştiente, atunci trebuie să fie ceva foarte subtil, cu siguranţă dincolo de calculabilitate, dincolo de haos, de asemenea, dincolo de influenţele întâmplătoare. Dacă un astfel de concept de 'cauză' poate să ne conducă mai aproape de înţelegerea problemei (sau iluziei) liberului arbitru, este o chestiune care îşi aşteaptă încă rezolvarea.

1.12. 'Conştienţă', 'înţelegere', 'conştiinţă', 'inteligenţă'

În cele discutate mai sus nu am făcut, deocamdată, nici un efort să fiu precis cu termenii derutanţi legaţi de cuvântul minte[10]. Despre diferitele calităţi ale minţii nu am vorbit până în acest punct, deşi m-am referit, destul de vag, în definiţiile date în paragraful 1.3 (punctele de vedere $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ şi $\mathcal{D}$), la noţiunea

[10] Am ales traducerea *mind* = *minte* deoarece este cea care pare să se fi impus în domeniul psihologiei şi al 'filosofiei minţii' în România. *Minte* este un termen mai larg răspândit decât intelect, raţiune, conştiinţă, etc. Termenul pe care unele traduceri îl mai folosesc este 'mental' (substantivat şi, uneori, articulat) (n.t.)

de conştienţă[11] Voi încerca acum să clarific cel puţin terminologia pe care o voi folosi în continuare, în special partea legată de termeni cum sunt înţelegere (*understanding*), conştiinţă (*consciousness*) şi inteligenţă (*intelligence*), termeni foarte importanţi pentru cele discutate în această carte.

Cum nu cred că încercarea de a defini complet exhaustiv aceşti termeni poate fi utilă, voi încerca să explic modul în care folosesc eu această terminologie, prin câteva comentarii. Adeseori sunt tulburat să observ că modul în care folosesc eu aceste cuvinte şi care mi se pare de la sine înţeles, nu concordă cu ceea ce altora li se pare natural. De exemplu, semnificaţia cuvântului *înţelegere* implică pentru mine şi faptul că simpla posedare a acestei calităţi impune prezenţa anumitor elemente de conştienţă.

Fără a fi conştient de ceea ce susţine o argumentaţie, nu poţi înţelege argumentaţia respectivă. Mie mi se pare că, fără excepţie, cuvintele se folosesc în acest mod, deşi în anumite contexte, susţinătorii **IA** par să folosească termenii *înţelegere* şi *conştienţă* într-un mod care neagă o astfel de implicaţie. Susţinătorii **IA** (atât cei ai lui $\mathcal{A}$, cât şi cei ai lui $\mathcal{B}$) afirmă că roboţii controlaţi de calculator 'înţeleg' instrucţiunile care li se dau, deşi nimeni nu afirmă că ei ar fi şi 'conştienţi' de semnificaţia acestora. Pentru mine acesta este un mod greşit de folosire a cuvântului *înţelegere*, o utilizare greşită, care are însă o valoare pur euristică pentru descrierea funcţionării calculatorului. Atunci când voi dori să fie clar că nu folosesc cuvântul *înţelegere* în acest mod euristic, voi spune 'înţelege în mod real' sau 'înţelegere reală', pentru activitatea în care conştienţa este necesară.

Bineînţeles, unii ar putea veni cu observaţia că nu există nici o distincţie clară între utilizarea cuvântului *înţelegere* în aceste două moduri. Dacă cineva crede că nu există nici o diferenţă, atunci trebuie să creadă de asemenea că termenul *conştienţă* este un concept prost definit. Eu nu neg acest lucru; dar *conştienţă* este de fapt *ceva* şi acest ceva poate fi prezent sau absent în grade diferite. Dacă cineva crede că ea, conştienţa, este 'ceva', atunci mi se pare normal ca el să fie de acord

[11] În limba engleză *awareness* şi *consciousness* sunt perfect sinonime. Totuşi, autorul le foloseşte aici cu sensuri diferite, bine definite. Am tradus *awareness* = *conştienţă* şi *consciousness* = *conştiinţă* , pentru motive ce vor deveni clare din paragraful de faţă. O utilizare similară a acestor două cuvinte apare în psihologia analitică, la C.C. Jung şi au fost traduse la fel în limba română. Vezi, de exemplu, Antologia C.C. Jung, apărută la Ed.Anima, Bucureşti 1994, volum tradus de Suzana Holan (n.t.)

că acel ceva face parte din orice înţelegere reală. Acest lucru permite ca acel *ceva* care este conştienţa să poată fi o caracteristică a unei activităţi pur calculative, în concordanţă cu punctul de vedere $\mathcal{A}$

Mi se pare de asemenea că utilizarea cuvântului *inteligenţă* trebuie făcută fără nici o excepţie, doar când se presupune că există deja *înţelegere*. Totuşi, din nou anumiţi susţinători **IA** ar putea afirma că roboţii lor ar putea fi 'inteligenţi' fără să trebuiască să 'înţeleagă' ceva. Termenul *inteligenţă artificială* implică faptul că se consideră posibilă o activitate calculatorie inteligentă, în vreme ce unii spun că înţelegerea reală (şi o anumită conştienţă) este în afara scopurilor **IA**. Din punctul meu de vedere, *inteligenţa* fără înţelegere înseamnă doar o utilizare greşită a cuvintelor. Până la un punct, o anumită simulare a inteligenţei pure, fără înţelegere, poate fi uneori posibilă (într-adevăr, nu foarte rar întâlneşti *oameni* care sunt capabili să ne păcălească pentru o vreme şi să ne facă să credem că posedă o anumită înţelegere, pentru ca în final să se vadă că de fapt nu înţeleg nimic!). Faptul că există o diferenţă clară între inteligenţa pură (sau înţelegerea pură) şi orice activitate simulată cu calculatorul va fi o idee importantă în cele ce urmează. Conform terminologiei mele, prezenţa unei inteligenţe pure impune şi prezenţa unei înţelegeri (pure). Deci, modul în care folosesc eu termenul *inteligenţă* (mai ales când va fi urmat de cuvântul *pur*) va implica prezenţa unei anumite conştienţe.

Consider această terminologie naturală, dar mulţi susţinători **IA** [16] (mai ales aceia care nu îmbrăţişează punctul de vedere $\mathcal{A}$) vor nega cu hotărâre că ei ar încerca să furnizeze o 'conştienţă' artificială, chiar dacă denumirea de **IA** implică faptul că încearcă să construiască o 'inteligenţă' artificială. Probabil că aceste persoane (în concordanţă cu punctul de vedere $\mathcal{B}$) vor să spună că doresc doar să *simuleze* inteligenţa — ceea ce nu necesită înţelegerea efectivă sau conştienţa — mai degrabă decât să atingă ceea ce eu numesc inteligenţă *pură*. Poate că ei vor pretinde că nu fac nici o diferenţă între inteligenţa pură şi inteligenţa simulată, aşa cum ar implica punctul de vedere $\mathcal{A}$ În argumentaţia pe care o voi face urmăresc să arăt că există într-adevăr un aspect al *inteligenţei pure* care nu poate fi simulat în nici un mod cu un calculator.

Evident, nu am definit nici unul dintre termenii *inteligenţă*, *înţelegere* sau *conştienţă*. Cred că ar fi foarte neînţelept să dau definiţii exhaustive aici. Va trebui să mă bazez într-o oarecare măsură pe intuiţia noastră despre ceea ce înseamnă de fapt aceste cuvinte. Dacă conceptul nostru intuitiv despre *înţelegere* este ceva de

care este nevoie pentru *inteligenţă*, atunci un argument care stabileşte natura non-computaţională a *înţelegerii* va stabili şi natura non-computaţională a *inteligenţei*. Mai mult, dacă *conştienţa* este ceva de care este nevoie pentru *înţelegere*, atunci o bază fizică non-computaţională pentru fenomenul conştienţei poate conta pentru a susţine natura non-computaţională a *înţelegerii*. Deci, propria mea utilizare a termenilor (şi, după cum susţin eu, utilizarea comună a termenilor) îndreptăţeşte implicaţiile

(a) *inteligenţa* cere *înţelegere*

şi

(b) *înţelegerea* cere *conştienţă*.

Conştienţa o consider ca un aspect — aspectul *pasiv* — al fenomenului *conştiinţă*. *Conştiinţa* are de asemenea şi un aspect activ şi anume, sentimentul liberului arbitru. Nu voi încerca aici o definiţie propriu zisă a cuvântului *conştiinţă* (şi cu atât mai puţin al liberului arbitru), chiar dacă scopul meu ultim este înţelegerea acestui fenomen al conştiinţei în termeni ştiinţifici, dar non-computaţionali — aşa cum cere punctul de vedere *C*

Fără a pretinde că aş fi făcut paşi importanţi pe drumul spre acest scop, totuşi sper că argumentele care le prezint în această carte (ca şi în MN) vor oferi câteva repere pe acest drum — şi poate chiar mai mult decât atât. Simt că încercarea de a defini cât mai precis termenul *conştiinţă* în acest stadiu ne-ar putea face să trecem pe lângă conceptul pe care tocmai vrem să-l lămurim. În consecinţă, în loc să încerc să dau o definiţie prematură şi inadecvată, voi face doar câteva comentarii descriptive privind modul în care eu folosesc termenul 'conştiinţă' După ce totul va fi spus şi făcut, trebuie, în cele din urmă, să ne bizuim pe intuiţia noastră asupra semnificaţiei termenului *conştiinţă*.

Asta nu înseamnă că intenţionez să sugerez că ştim cu adevărat, pe bază de intuiţie, ce *este* de fapt conştiinţa, ci pur şi simplu că există un astfel de concept pe care vrem să-l pătrundem într-un fel — un fenomen care poate fi descris pur ştiinţific, care joacă un rol activ, precum şi unul pasiv în lumea fizică. Unele persoane par să creadă că acest concept este prea vag pentru a merita să fie studiat. Mai mult, astfel de persoane [17] nu ezită să discute în schimb conceptul de *minte*, de parcă el ar fi mai bine definit. Folosirea uzuală a cuvântului *minte* presupune că el s-ar referi (şi chiar se referă) la ceva ce adeseori este considerat *inconştient*.

Pentru felul meu de a gândi, văd o mult mai mare obscuritate în legătură

cu conceptul de inconştient decât cu cel de minte conştientă. Deşi eu însumi folosesc cuvântul minte destul de rar, nu fac nici o încercare de a-l preciza. Conceptul de *minte* nu va ocupa un loc important în încercarea mea de a avea o discuţie riguroasă — în afară de ceea ce este deja inclus în termenul *conştiinţă*.

Deci, ce înţeleg prin conştiinţă ? Am remarcat mai sus că există un aspect activ şi unul pasiv al conştiinţei, dar nu totdeauna este clar că se face o distincţie între acestea. Percepţia culorii roşu, pe de o parte, necesită cu siguranţă conştiinţă pasivă, la fel ca şi senzaţia de durere sau de plăcere pe care o trezeşte o melodie. Conştiinţa activă este implicată în actul de voinţă de a te scula din pat dimineaţa, sau în cazul deciziei deliberate de a înceta o activitate consumatoare de energie.

Aducerea în minte a unor amintiri vechi implică atât aspectele active cât şi cele pasive ale conştiinţei. Conştiinţa, activă şi pasivă, va fi în mod normal implicată în formularea unui plan de acţiune viitor şi este cu siguranţă un ingredient necesar în acel tip de activitate mentală care este în mod obişnuit denumită *înţelegere*. Mai mult, putem fi conştienţi (pasiv) într-o anumită măsură chiar atunci când suntem adormiţi, cu condiţia să visăm (şi partea activă a conştiinţei să înceapă să joace un rol, îndată ce ne trezim).

Unii ar putea să conteste implicarea unui singur concept, cel de conştiinţă, în toate aceste manifestări. Ei ar putea insista că există şi alte atribute ale conştiinţei decât cele de *activ* şi *pasiv*. Ca atare, aplicarea, ca o umbrelă, a termenului pentru toate acele lucruri care ar putea fi considerate ca atare, ar fi, în cel mai bun caz inutilă. În concepţia mea există cu adevărat conceptul unificator *conştiinţă* care este central pentru toate aspectele separate ale activităţii mentale. Deşi admit că există aspecte pasive şi active ale conştiinţei care pot fi distinse, cele pasive fiind legate de senzaţii (sau *qualia*) iar cele active de liberul arbitru, iau aceste două aspecte ca pe cele două feţe ale aceleiaşi medalii.

Mă voi ocupa, în prima parte a acestei cărţi de ceea ce putem obţine utilizând activitatea mentală numită *înţelegere*. Deşi nu intenţionez să definesc ce înseamnă acest cuvânt, sper că sensul său este clar, astfel încât cititorul să fie convins că această activitate trebuie să fie o parte esenţială pentru acele calităţi care sunt necesare pentru a accepta argumentele din paragraful 2.5. Îmi propun să arăt că aprecierea *acestor* argumente trebuie să includă ceva non-computaţional. Argumentaţia mea nu este *direct* legată de termenii *inteligenţă*, *conştienţă*, *conştiinţă*, sau *minte*, dar ar trebui ca discuţia sa aducă clarificări serioase privind

aceste concepte, deoarece terminologia de bun simţ, pe care am indicat-o anterior, obligă conştienţa să fie un ingredient al înţelegerii, iar înţelegerea să fie parte a inteligenţei pure.

1.13. Argumentul lui John Searle

Înainte de a prezenta propria mea concepţie, mă voi referi pe scurt la studiile filosofului John Searle [18] — este vorba despre bine cunoscuta 'Cameră chinezească' — pentru a sublinia caracterul deosebit şi intenţiile de bază ale propriului meu raţionament. Analiza lui Searle se ocupă tot de *înţelegere* şi încearcă să stabilească dacă despre acţiunea unui calculator destul de sofisticat se poate spune că atinge această calitate mentală. Nu voi repeta dezvoltarea lui Searle în detaliu, ci voi reda pe scurt esenţialul.

Este vorba despre un program de calculator care-şi propune să simuleze 'înţelegerea' dând răspunsuri la întrebările puse despre o poveste care îi fusese spusă mai înainte — toate întrebările şi răspunsurile fiind în limba chineză. Apoi Searle propune unui colaborator, ignorant în limba chineză, să muncească serios, astfel încât să execute în detaliu toate calculele pe care calculatorul le-a făcut pentru a da răspunsurile. În ciuda aparenţei de înţelegere care rezultă din răspunsurile calculatorului, colaboratorul, care a executat aceleaşi calcule, nu a înţeles nimic din poveste deşi a reuşit să reproducă răspusurile calculatorului.

Searle comentează, în concluzie, că înţelegerea, ca o calitate mentală, nu poate fi doar o chestiune de calcul — deoarece subiectul uman (care nu înţelegea limba chineză) putea să obţină aceleaşi răspunsuri ca ale calculatorului, fără să înţeleagă absolut nimic din poveste. Searle acceptă că o *simulare* a înţelegerii ar fi posibilă, în conformitate cu punctul de vedere $\mathcal{B}$, căci el admite că simularea poate fi realizată de un calculator care execută fiecare acţiune fizică a creierului (făcând ceea ce face acesta), atunci când înţelege cu adevărat ceva.

Folosind exemplul 'Camerei chinezeşti', Searle insistă că o simulare nu poate, prin ea însăşi, să 'simtă' că înţelege ceva. Astfel, înţelegerea *de fapt* nu poate fi obţinută prin simulare pe calculator.

Exemplul lui Searle este îndreptat direct contra lui $\mathcal{A}$ (care susţine că orice *simulare* a înţelegerii ar fi echivalentă cu înţelegerea *de fapt*) şi este prezentat

ca fiind în sprijinul lui $\mathcal{B}$ (şi, în acelaşi mod, un suport pentru $\mathcal{C}$ şi $\mathcal{D}$). Exemplul priveşte aspectele pasive, interioare sau subiective ale înţelegerii. El nu neagă însă posibilitatea simulării înţelegerii în aspectele sale active, exterioare şi obiective. Chiar Searle însuşi este înregistrat pe bandă magnetică spunând

> *Evident, creierul este un calculator digital. Cum totul este un calculator digital, creierul este şi el unul.* [19]

Acest lucru sugerează că el ar putea fi de acord să accepte posibilitatea unei simulări complete a activităţii creierului conştient în actul *înţelegerii*, atâta vreme cât manifestările exterioare ale acestei simulări ar fi identice cu cele ale unei fiinţe umane reale — în concordanţă cu punctul de vedere $\mathcal{B}$.

Pe de altă parte, argumentele mele vor fi îndreptate împotriva acestor aspecte de 'înţelegere' aparentă, voi susţine că nici măcar o simulare adecvată pe calculator a manifestărilor exterioare ale înţelegerii nu este posibilă. Nu mai prezint alte detalii ale discuţiei lui Searle deoarece ea nu oferă nici o susţinere directă a punctului de vedere $\mathcal{C}$ (susţinerea lui $\mathcal{C}$ este scopul meu în cele ce urmează).

Totuşi, vreau să consemnez că privesc 'Camera chinezească' drept un exemplu convingător împotriva lui $\mathcal{A}$, deşi nu cred că este pe de-a-ntregul concludent.

Pentru alte detalii şi diferite contraexemple, vezi Searle (1980) şi discuţiile din cartea editată de Hofstadter şi Dennett (1981); vezi de asemenea Dennett (1990) şi Searle (1992). Pentru evaluarea mea proprie vezi MN p. 26-32.

1.14. Unele dificultăţi ale modelului computaţional

Înainte de a ne întoarce la ceea ce separă punctul de vedere $\mathcal{C}$ de $\mathcal{A}$ şi $\mathcal{B}$, să considerăm câteva dintre dificultăţile cărora trebuie să le facă faţă orice încercare de explicare a fenomenului conştiinţei în concordanţă cu punctul de vedere $\mathcal{A}$. În conformitate cu $\mathcal{A}$, efectuarea unui algoritm (adică activarea lui), este cea despre care se presupune că trezeşte conştienţa. Dar ce înseamnă asta de fapt? Activarea algoritmului înseamnă că bucăţi din materialul fizic trebuie mişcate în conformitate cu operaţiile succesive ale algoritmului ? Să ne imaginăm aceste instrucţiuni scrise linie cu linie într-o carte mare [20]. Acţiunea de scriere sau tipărire a acestor linii constituie o activare? Simpla existenţă statică a cărţii ar fi

suficientă? Ce se poate spune în cazul în care un deget se mişcă urmărind liniile una după alta; ar putea conta asta ca activare? Dar dacă degetul ar urmări liniile scrise în alfabet Braille ? Dar dacă aceste pagini ar fi proiectate pe un ecran? Este simpla prezentare a instrucţiunilor succesive ale unui algoritm o activare? Ar putea fi necesar, pe de altă parte, să avem pe cineva care să verifice dacă fiecare linie urmează celeilalte în conformitate cu regulile algoritmului în cauză? Poate că acest lucru ar depăşi problema, deoarece înţelegerea unei persoane auxiliare (conştiente) nu este necesară în proces. Întrebarea: „Ce acţiuni fizice alcătuiesc un algoritm în proces de efectuare?" este profund neclară. Probabil că astfel de acţiuni nu sunt necesare deloc şi, pentru a fi în concordanţă cu $\mathcal{A}$, chiar simpla existenţă matematică 'platonică' a algoritmului (în conformitate cu 1.7) ar fi suficientă pentru ca să fie prezentă *conştienţa*.

În orice caz, nici măcar în conformitate cu $\mathcal{A}$, nu este posibil cazul ca fiecare algoritm complicat să trezească conştienţa (într-o cantitate apreciabilă). Ar fi de aşteptat ca anumite caracteristici speciale ale algoritmului, ca de exemplu 'universalitatea', auto-referenţierea', sau complexitatea algoritmică [21], sau altele de acelaşi fel, să fie necesare pentru ca să putem considera că apare o 'cantitate semnificativă' de conştienţă. Mai mult, este stânjenitor să stabileşti ce calităţi speciale ale unui algoritm se presupune că sunt responsabile pentru diferitele calităţi care constituie conştienţa noastră. Ce fel de calcul evocă senzaţia de 'roşu', de exemplu ? Ce calcule constituie senzaţia de 'durere', 'duioşie', 'armonie', 'picant' sau altele ? S-au făcut încercări de către susţinătorii lui $\mathcal{A}$ pentru a aborda probleme de această natură (vezi, de exemplu, Dennett 1991), dar până acum aceste încercări nu mi s-au părut deloc convingătoare.

Mai mult, orice sugestie bine formulată şi suficient de simplă algoritmic (aşa cum nu este nici una în literatura de până acum), ar avea de suferit, din cauză că ar putea fi implementată fără mare dificultate pe un calculator din zilele noastre. O astfel de implementare, spun susţinătorii unei astfel de sugestii, trebuie să provoace experienţa *actuală* a 'qualium'-ului intenţional.

Este greu, chiar şi pentru aceia care aderă cu convingere la punctul de vedere $\mathcal{A}$, să accepte cu seriozitate că un astfel de calcul — sau orice calcul care poate fi executat pe calculatoarele de astăzi, utilizând cunoştinţele de acum din **IA** — ar putea cu adevărat să experimenteze activitatea mentală la un nivel semnificativ. S-ar putea ca susţinătorii unor astfel de idei să creadă că extraordinara

complicaţie a calculelor (acţionând în conformitate cu aceste sugestii) care sunt implicate în activitatea creierului ne permit să avem o experienţă mentală apreciabilă.

Acest lucru ridică alte întrebări pe care nu le-am văzut analizate încă la un nivel semnificativ.

Dacă complicaţia extraordinară a reţelei neuronale şi sinaptice din creier este considerată responsabilă pentru activitatea mentală conştientă, atunci ar trebui să ne obişnuim cu ideea că nu toate părţile creierului uman au de-a-face cu conştiinţa. Atunci când se vorbeste despre *creier* fără altă calificare, este natural (cel puţin pentru nespecialişti) să ne gândim la regiunile pline de convoluţii care constituie ceea ce se numeşte cortexul cerebral — materia cenuşie exterioară a creierului (cerebrum).

Există aproximativ o sută de mii de milioane (10^{11}) neuroni în cortexul cerebral — care, într-adevăr, implică mari complicaţii — dar cortexul cerebral nu reprezintă, nici pe departe, tot creierul. Sub el şi în spate se află o altă parte importantă — o masă de neuroni cunoscută sub numele de cerebel (figura 1.6).

Cerebelul este implicat în mod hotărâtor în comanda mişcărilor (controlul motor) şi joacă un rol important în anumite deprinderi motorii care sunt dirijate aproape mecanic, nu prin gândire conştientă — devenind 'o a doua natură' Iniţial, atunci când anumite deprinderi sunt învăţate, controlul conştient al acţiunilor este necesar şi se obţine prin implicarea esenţială a cortexului cerebral. Apoi, după ce mişcările au devenit 'automate', sunt preluate de activitatea inconştientă a cerebelului.

Trebuie remarcat, în lumina faptului că activitatea cerebelului este în întregime inconştientă, că el implică cam pe jumătate atâţia neuroni ca cerebrum-ul. Mai mult, celulele lui Purchinje, despre care am vorbit în paragraful 1.2, care au până la 80000 de legături sinaptice, sunt neuroni care se află în cerebel, astfel numărul total de legături între neuroni ar putea să nu fie mai mic în cerebel decât în cerebrum. Dacă se consideră că o reţea extraordinar de complicată este cerinţa esenţială pentru conştiinţă, ne putem întreba de ce conştiinţa pare să lipsească cu totul în activitatea cerebelului (voi avea de făcut câteva comentarii în legătură cu această problemă în paragraful 8.6).

Bineînţeles, problemele legate de punctul de vedere $\mathcal{A}$, despre care am vorbit în această secţiune, îşi au corespondentele pentru $\mathcal{B}$ şi $\mathcal{C}$. În orice demers

ştiinţific trebuie să ne punem întrebarea ce anume se află la baza fenomenului conştiinţei şi cum poate fi el caracterizat.

În secţiunile părţii a doua, voi încerca să mă îndrept spre o înţelegere a conştiinţei din punctul de vedere al poziţiei $\mathcal{C}$

1.15. Limitările actuale din IA oferă sprijin poziţiei $\mathcal{C}$?

Dar de ce tocmai $\mathcal{C}$? Există vreun motiv care s-ar putea interpreta ca suport direct pentru $\mathcal{C}$? Este $\mathcal{C}$ o alternativă serioasă la $\mathcal{A}$ sau $\mathcal{B}$ sau chiar $\mathcal{D}$? Trebuie să încercăm să vedem ce putem face, de fapt, cu mintea noastră (cu creierul nostru) atunci când intervine în joc analiza conştientă — şi voi încerca să conving cititorul că, (cel puţin uneori) ceea ce facem atunci când gândim conştient, este foarte diferit de tot ceea ce se poate obţine prin calcul.

Aderenţii opiniei $\mathcal{A}$ vor dori să susţină că, într-o formă sau alta, 'calculul' este singura explicaţie — şi, cel puţin în privinţa efectelor asupra comportării exterioare, la fel vor spune şi susţinătorii punctului de vedere $\mathcal{B}$ Pe de altă parte, aderenţii la $\mathcal{D}$, ar putea să fie de acord cu $\mathcal{C}$, atunci când cred că acţiunile conştiente trebuie să fie ceva situat deasupra calculului, dar ei neagă posibilitatea unei explicaţii, în termeni ştiinţifici, a conştiinţei. Astfel, pentru a oferi suport opiniei $\mathcal{C}$, trebuie, pe de o parte, găsite exemple de activitate mentală care să fie deasupra oricărei forme de calcul, iar pe de altă parte trebuie văzut dacă această activitate poate rezulta din procese fizice specifice. Partea întâi din această carte este dedicată primului dintre aceste scopuri, iar partea a doua celui de-al doilea.

Ce fel de activitate mentală ar fi aceea pe care să o considerăm că se află în afara calculului? O variantă ar fi să încercăm să examinăm starea prezentă a inteligenţei artificiale şi să încercăm să vedem la ce sunt bune sistemele comandate prin calcul şi la ce nu sunt bune. Evident că starea actuală a IA nu ne poate da o indicaţie clară despre ceea ce se poate realiza în principiu Dar, în 50 de ani, să zicem, lucrurile vor arăta foarte diferit de cele din prezent. Dezvoltarea rapidă a calculatoarelor şi a aplicaţiilor lor — numai în ultimii 50 de ani — a fost extraordinară.

Trebuie să fim pregătiţi pentru progrese enorme în viitor — progrese care ar putea veni peste noi foarte rapid. În această carte mă voi ocupa mai puţin de viteza acestor progrese şi mai mult de limitările fundamentale de principiu la care sunt supuse. Aceste limitări se vor aplica într-un viitor care se poate întinde pe secole. De aceea, ne vom baza argumentele pe principii generale şi nu ne vom lăsa influenţaţi de succesele de moment. Fără îndoială, pot fi găsite lucruri interesante în succesele şi eşecurile inteligenţei artificiale actuale, în ciuda faptului că, până acum, prea puţin din ceea ce s-ar putea numi inteligenţă artificială există — aşa cum trebuie să admită şi cei mai înfocaţi susţinători ai **IA**

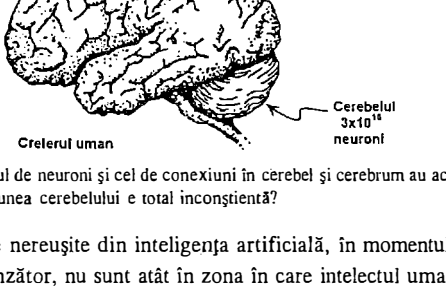


Fig. 1.6. Numărul de neuroni şi cel de conexiuni în cerebel şi cerebrum au acelaşi ordin de mărime. Atunci de ce acţiunea cerebelului e total inconştientă?

Marile nereuşite din inteligenţa artificială, în momentul de faţă, poate nu în mod surprinzător, nu sunt atât în zona în care intelectul uman poate fi el însuşi spectaculos — ca acolo unde experţii ne uimesc cu cunoştinţele lor de specialitate sau cu capacitatea lor de a face raţionamente bazaţi pe proceduri de calcul complicate — ci în activităţile 'de bun simţ' , pe care cel mai umil dintre noi le îndeplineşte de obicei în activitatea de zi cu zi.

Deocamdată, nici un robot comandat de un calculator nu poate concura cu un copil mic atunci când trebuie executate unele activităţi simple cu ar fi să recunoască un creion colorat aflat pe jos în celălalt colţ al camerei, care îi este necesar pentru a completa desenul, să traverseze camera, să ridice creionul şi apoi să-l folosească la colorat. La fel, capacităţile unei furnici în îndeplinirea activităţilor sale zilnice, depăşesc cu mult ceea ce poate realiza cel mai sofisticat sistem comandat de un calculator din zilele noastre. Dar, pe de altă parte, dezvoltarea calculatoarelor puternice de jucat şah ne oferă un exemplu strălucit în care calculatorul poate fi extrem de eficient. Şahul este, fără îndoială, locul unde

puterea intelectului uman se manifestă în mod special — exploatată strălucit în acest mod, dar totuşi destul de puţin. Deja sistemele de calcul proiectate pentru jocul de şah joacă extrem de bine şi îi pot bate, aproape sigur, pe cei mai mulţi jucători de şah. Chiar cei mai mari maeştri sunt acum puternic presaţi, căci, nu peste multă vreme[12], îşi vor pierde superioritatea pe care încă o au asupra celui mai bun calculator [22]. Mai există şi alte domenii în care calculatoarele îi pot concura cu succes, sau măcar cu un succes parţial, pe experţii umani. Mai mult, există domenii, cum este calculul numeric, în care posibilităţile calculatoarelor depăşesc foarte mult capacitatea umană.

În toate aceste cazuri, este greu, totuşi, să susţii despre calculator că atinge vreo înţelegere reală în legătură cu ceea ce are de făcut. În cazul organizării *top-down*, aceste sisteme lucrează cu succes, nu pentru că ele înţeleg ceva, ci pentru că înţelegerea programatorilor (sau înţelegerea experţilor de care depindeau programatorii) a fost utilizată în construirea programului. Pentru organizarea *bottom-up* nu este clar că ar fi nevoie de o înţelegere, ca o caracteristică a acţiunilor sistemului, atât din partea dispozitivului propriu zis, cât şi a programatorilor săi — în afara acelei înţelegeri de care este nevoie în proiectarea detaliilor algoritmilor care se perfecţionează şi a concepţiei că sistemul poate să-şi îmbunătăţească performanţele pe baza experienţei ori de câte ori are încorporat un sistem *feedback*. Evident, nu totdeauna este limpede ce înseamnă de fapt termenul *înţelegere*, aşa încât unii ar putea susţine că în sensul lor, aceste sisteme de calcul chiar au un anume tip de *înţelegere*.

Dar este acest lucru raţional ? Pentru a ilustra lipsa oricărei înţelegeri la calculatoarele actuale este interesant să dăm un exemplu, poziţia de pe tabla de şah din figura 1.7 (datorată lui William Hartson şi luată dintr-un articol de Jeane Seymore şi David Norwood (1993)). În această poziţie, negrul are un avantaj material enorm, având două ture şi un nebun în plus. Totuşi, este simplu pentru alb să evite înfrângerea dacă mişcă tot timpul regele în partea sa de tablă. Zidul pionilor este inexpugnabil pentru piesele negre şi deci turele negre şi nebunul nu reprezintă nici un pericol pentru alb. Această partidă este clară pentru orice jucător familiarizat cât de cât cu regulile şahului. Totuşi, când poziţia, cu albul la mutare,

---

[12] Partida 'decisivă' dintre campionul mondial Gari Kasparov şi calculatorul numit Deep Blue s-a desfăşurat la 4 ani după finalizarea acestei cărţi ! (n.t.)

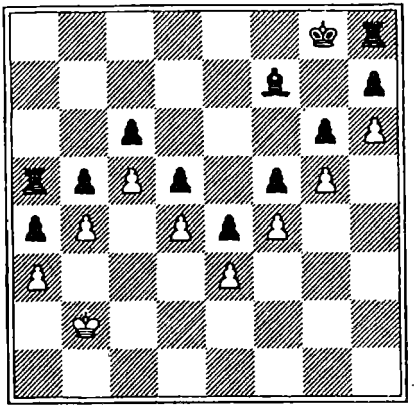


Fig. 1.7. Albul la mutare - o decizie uşoară pentru oameni, dar Deep Thought ia tura !

a fost prezentată lui *Deep Thought* — cel mai puternic calculator de jucat şah din acel moment, având în palmares multe victorii asupra unor mari maeştri (oameni) — acesta s-a repezit 'orbeşte' să ia tura neagră cu pionul deschizând astfel bariera pionilor negri, pentru a ajunge într-o poziţie fără speranţă.

Cum a putut un jucător de şah atât de experimentat să facă o mutare atât de stupidă ? Răspunsul este că *Deep Thought* fusese programat, în plus faţă de 'cartea de şah' de dimensiune mare pe care o memorase, să calculeze mutare după mutare — cu o considerabilă profunzime — şi să încerce să-şi îmbunătăţească raportul de piese de pe tablă. În nici un moment nu avusese nici o înţelegere legată de ceea ce înseamnă o barieră de pioni — nici vreo altă înţelegere în legătură cu ceea ce făcea.

Pentru cel ce este suficient familiarizat cu modul (general) în care Deep Thought, sau alt calculator — jucător de şah, este construit, nu este o surpriză reală că el a eşuat într-o poziţie ca cea din figura 1.7 Nu numai că noi putem înţelege câte ceva despre şahul pe care *Deep Thought* nu-l înţelege, dar noi putem înţelege de asemenea câte ceva despre procedura (*top-down*) pe baza căreia *Deep Thought* a fost construit; astfel, putem aprecia şi de ce a putut face o asemenea gafă — precum şi să înţelegem de ce poate juca şah atât de bine în cele mai multe împrejurări. Totuşi ne putem întreba este oare posibil ca *Deep Thought*, sau ori care alt sistem **IA**, să poată ajunge la vreo înţelegere reală, de felul celei pe care

noi o avem — la şah, sau în orice altă împrejurare ?

Unii susţinători ai IA ar putea argumenta că, pentru ca un sistem IA să câştige o capacitate de înţelegere 'reală', ar trebui să fie programat pe baza unor proceduri *bottom-up*, în măsură mult mai mare decât sunt programate în momentul de faţă sistemele specializate în jocul de şah. Astfel, capacităţile lor de 'înţelegere' s-ar dezvolta progresiv prin utilizarea 'experienţei' acumulate, mai degrabă decât dacă ar avea implementate reguli algoritmice *top-down*. Regulile *top-down*, care sunt simple pentru noi, nu pot, prin ele însele, să ofere o bază pentru o înţelegere reală — deoarece noi putem folosi chiar înţelegerea acestor reguli pentru a realiza limitările lor fundamentale.

Aceste idei vor fi reluate mai explicit în prezentarea făcută în capitolele 2 şi 3. Dar ce este de fapt cu aceste proceduri de calcul numite *bottom-up* ? Este oare posibil ca ele să furnizeze o bază pentru înţelegere? În capitolul 3 voi demonstra contrariul. Pentru moment, trebuie pur şi simplu să luăm act de faptul că sistemele de calcul din zilele noastre nu pot înlocui, în nici un fel, capacitatea umană de înţelegere pură — în nici un domeniu de competenţă intelectuală unde înţelegerea umană permanentă şi intuiţia sunt esenţiale. Sunt sigur că aceste lucruri sunt astăzi general acceptate. În cea mai mare parte promisiunile [23] foarte optimiste care au fost cândva făcute de susţinătorii înteligenţei artificiale şi promotorii sistemelor expert nu au fost încă îndeplinite. Dar suntem încă foarte la început, dacă ar fi să luăm în considerare ceea ce inteligenţa artificială ar putea în cele din urmă obţine. Susţinătorii IA (atât cei ai lui $\mathcal{A}$ cât şi cei ai lui $\mathcal{B}$) vor susţine că este doar o problemă de timp şi, probabil, de anumite dezvoltări semnificative ale meseriei, înainte ca elemente importante ale înţelegerii să înceapă să fie vizibile în comportamentul sistemelor lor, comandate de calculator. În cele ce urmează voi încerca să vin cu argumente precise împtriva acestei opinii şi să arăt că există limitări fundamentale pentru orice sistem pur computaţional, oricum ar fi el, *top-down* sau *bottom-up*. Deşi s-ar putea să fie posibil ca un astfel de sistem construit cu suficientă dibăcie (ca şi Deep Thought) să ofere o iluzie, pentru o bucată de vreme, că posedă o anumită înţelegere, voi susţine că lipsa actuală de înţelegere a unui sistem de calcul se va menţine.

Pentru argumentaţia mea, va trebui să mă întorc din nou la matematică, intenţia fiind să arăt că înţelegera matematică este ceva ce nu se poate reduce la calcul. Unii dintre susţinătorii IA ar putea găsi acest lucru surprinzător deoarece

ei au susţinut [24] că acele lucruri care au fost obţinute mai târziu în evoluţia umanităţii, ca de exemplu, calculul aritmetic sau algebric, sunt lucrurile care vor putea fi puse mai întâi pe calculator — şi aici calculatoarele au depăşit deja spectaculos capacitatea de calcul a oamenilor — în vreme ce acele deprinderi care apar mai devreme, ca umblatul sau interpretarea vizuală a unor scene, sunt lucruri pe care noi le facem fără efort, în vreme ce calculatoarele actuale se luptă să atingă nişte performanţe foarte modeste. Voi aborda însă altfel problema. Orice activitate complicată, cum ar fi calculul matematic, jocul de şah, sau alte activităţi comune — *dacă* au fost înţelese în termenii unor reguli clare de calcul — sunt lucruri pe care calculatoarele moderne le pot face bine; dar chiar înţelegerea, care stă la baza regulilor de calcul, este un lucru care el, în sine, stă deasupra calculului.

1.16. Argumentul oferit de teorema lui Gödel

Cum putem fi siguri că înţelegerea nu este ea însăşi un lucru care s-ar putea reduce la reguli de calcul? Voi da pe scurt (în capitolele 2 şi 3) câteva motive foarte serioase de a vedea că efectele înţelegerii nu pot fi simulate corespunzător în nici un fel de termeni computaţionali — nici utilizând o organizare *top-down* nici *bottom-up* şi nici cu vreo combinaţie a acestora două. Astfel, capacitatea umană de a înţelege este ceva ce trebuie atins printr-o activitate non-computaţională a minţii sau creierului.

Poate că cititorul îşi aminteşte că (aşa cum spuneam în 1.5, 1.9) termenul 'non-computaţional' se referă la ceva ce se situează dincolo de orice tip de simulare efectivă cu ajutorul unui calculator bazat pe principiile logicii care stau la baza tuturor dispozitivelor de calcul electronice sau mecanice din zilele noastre. Pe de altă parte, 'activitatea non-computaţională' nu implică nimic care să se afle mai presus de posibilităţile ştiinţei şi ale matematicii. Dar ea implică faptul că punctele de vedere $\mathcal{A}$ şi $\mathcal{B}$ nu pot explica modul în care realizăm de fapt toate acele sarcini care sunt rezultatul activităţii mentale conştiente.

Cu siguranţă există o posibilitate *logică*, ca mintea conştientă să acţioneze după astfel de legi non-computaţionale (conform paragrafului 1.9). Dar este oare adevărat ?

Argumentaţia pe care o voi prezenta în capitolul următor (2.5) va oferi

ceea ce eu cred că este o argumentaţie bine construită despre ingredientul non-computaţional al gândirii noastre conştiente. Aceasta depinde de o formă simplă a teoremei celebre şi puternice, datorată logicianului de origine cehă, Kurt Gödel. Voi prezenta o formă simplificată a demonstraţiei sale, care cere foarte puţină matematică (pentru aceasta voi împrumuta o idee mai târzie datorată lui Alan Turing). Oricare cititor bine intenţionat şi interesat de problemă sper să accepte că nu este prea greu să o urmărească. Totuşi, demonstraţia de tip Gödel utilizată în acest mod a fost uneori puternic criticată [25]. Ca urmare unii dintre cititori s-ar fi putut să rămână cu impresia că această argumentaţie, utilizând teorema lui Gödel, a fost respinsă în întregime. Voi arăta cât se poate de clar că lucrurile nu stau aşa. Este adevărat că s-au adus de-a lungul anilor multe contra-argumente. Multe dintre acestea se datorează unei abordări de pionierat — în favoarea mentalismului şi opusă fizicalismului — care a fost dezvoltată de filozoful John Lucas, de la Oxford (1961).

Lucas a argumentat, pornind de la teorema lui Gödel, că facultăţile mentale trebuie să se situeze deasupra a ceea ce este computaţional (alţii, ca Nagel şi Newman(1958), erau cam de aceeaşi părere). Argumentaţia mea, deşi urmează cam aceeaşi linie, este prezentată oarecum diferit de cea a lui Lucas — şi nu pentru că s-ar vrea un suport pentru mentalism. Eu cred că forma în care prezint lucrurile este mai bine pregătită să facă faţă la diferitele critici care s-au ridicat la adresa lui Lucas.

La momentul potrivit (capitolele 2 şi 3), mă voi referi, în amănunt, la toate contra-argumentele care mi-au reţinut atenţia. Eu sper că modul în care tratez subiectul va servi să corecteze, nu numai unele prejudecăţi larg răspândite privind semnificaţia demonstraţiei lui Gödel, ci şi să îmbunătăţească argumentaţia mea, poate cam superficială, dată în MN.

Voi demonstra că o mare parte dintre contra-argumente se bazează în principal pe prejudecăţi; celelalte, care aduc puncte de vedere originale care trebuie analizate în detaliu, pot să furnizeze doar posibilităţi pentru $\mathcal{A}$ şi $\mathcal{B}$, dar aşa cum voi arăta, ele nu furnizează explicaţii plauzibile despre ceea ce capacitatea noastră de înţelegere ne permite să obţinem şi, în orice caz, ele au prea puţină valoare pentru IA.

Orice susţinător al ideii după care manifestările exterioare ale procesului de gândire conştientă pot fi propriu zis simulate computaţional, în concordanţă cu

punctele de vedere $\mathcal{A}$ şi $\mathcal{B}$, trebuie să găsească un mod în care să facă faţă argumentelor pe care le voi da.

1.17. Platonism sau misticism ?

Unii critici ar putea totuşi să spună că, obligându-ne, în aparenţă, să aderăm la punctul de vedere $\mathcal{C}$ sau $\mathcal{D}$, raţionamentul lui Gödel are implicaţii care ar putea fi considerate 'mistice' şi, cu siguranţă, cu nimic mai atrăgătoare decât oricare altă posibilitate de a evita rezultatul lui Gödel. În ceea ce priveşte $\mathcal{D}$, sunt de fapt de acord cu aceştia. Motivele pentru care resping punctul de vedere $\mathcal{D}$ — cel care susţine că ştiinţa este neputincioasă să explice mintea umană — provin din faptul că doar utilizarea metodelor ştiinţei şi matematicii au condus la progrese reale în înţelegerea lumii. Mai mult, singurele manifestări mentale, despre care avem cunoştinţe directe, sunt cele intim asociate acelui obiect fizic numit creier — iar diferenţele în stările minţii par să fie, în mod clar, asociate cu diferenţe în stările fizice ale creierului. Chiar starea mentală numită conştiinţă pare să fie asociată cu anumite tipuri de activităţi fizice care au loc în creier. Dacă n-ar fi aspectele încurcate ale conştiinţei, legate de prezenţa *conştienţei* şi, probabil, de sentimentul liberului arbitru, care, deocamdată, par să scape descrierii fizice, n-ar fi necesar să ne simţim tentaţi să privim dincolo de metodele standard ale ştiinţei pentru explicarea minţii ca trăsătură a comportamentului fizic al creierului.

Pe de altă parte, ar trebui să fie limpede că ştiinţa şi matematica au descoperit o lume plină de mistere. Cu cât devine mai adâncă înţelegerea noastră ştiinţifică, cu atât este mai profund misterul revelat. Este de notorietate că fizica, cea mai familiarizată cu modul încurcat şi misterios în care se comportă de fapt materia, tinde să aibă o concepţie mai puţin clasic mecanicistă, decât au biologii. În capitolul 5 voi explica câteva dintre cele mai misterioase aspecte al comportării cuantice, unele dintre ele descoperite de curând. S-ar putea să avem nevoie, pentru a ne acomoda misterului minţii, de o lărgire a ceea ce în prezent înseamnă *ştiinţă*, dar eu nu văd nici un motiv să ne rupem de acele metode care ne-au servit atât de strălucit până acum. Dacă, aşa cum cred eu, raţionamentul lui Gödel ne forţează, prin consecinţă, la acceptarea punctului de vedere $\mathcal{C}$, într-o anumită formă, atunci va trebui să acceptăm şi alte implicaţii.

Suntem conduşi spre un punct de vedere platonic. În conformitate cu teoria lui Platon, conceptele şi adevărurile matematice există într-o lume a lor proprie, fără timp şi fără loc fizic. Lumea lui Platon este o lume ideală de forme perfecte, distinctă de lumea fizică, dar în termenii căreia lumea fizică poate fi înţeleasă. Ea, lumea lui Platon, se află deasupra construcţiilor noastre mentale imperfecte; totuşi, mintea noastră are un anumit acces la acest domeniu prin conştientizarea formelor matematice şi prin capacitatea noastră de a raţiona cu ele. Vom descoperi că, în timp ce percepţiile noastre platonice pot fi ajutate ocazional de calcul, ele nu sunt limitate de acesta. Acest potenţial al *conştienţei* conceptelor matematice, implicat în accesul platonic, este cel care dă minţii o putere mult mai mare decât cea care ar putea fi atinsă de un dispozitiv depinzând doar de calcul pentru acţiunile sale.

1.18. Cât de relevantă este înţelegerea matematică?

Unii dintre dumneavoastră s-ar putea plânge că toate acestea sunt doar speculaţii. Ce relevanţă au problemele sofisticate ale matematicii şi ale filosofiei matematicii pentru cele mai multe dintre chestiunile de interes direct din inteligenţa artificială, de exemplu? Într-adevăr, mulţi filosofi şi suporteri IA sunt cu totul de acord că, deşi în contextul originar al logicii matematice, teorema lui Gödel este, fără îndoială, importantă, ea poate avea implicaţii foarte limitate, în cazul cel mai bun, pentru IA sau pentru filozofia minţii. Foarte puţin din activitatea mentală a omului este îndreptată către chestiunile legate de contextul originar al teoremei lui Gödel: fundamentele axiomatice ale matematicii. Răspunsul meu este că o mare parte a activităţii mentale a omului implică, pe de altă parte, aplicarea conştiinţei umane şi a înţelegerii. Am utilizat raţionamentul lui Gödel pentru a arăta că înţelegerea umană nu poate fi o activitate algoritmică. Dacă pot arăta aceasta într-un context specific, acest lucru va fi suficient. Odată arătat faptul că anumite tipuri de înţelegere matematică pot eluda descrierea algoritmică, atunci este demonstrat că putem face ceva cu mintea noastră care să nu fie calculabil. Acest lucru fiind acceptat, este un pas natural să conchidem că activitatea necalculabilă trebuie să fie prezentă în multe alte aspecte ale activităţii mentale. Stăvilarul va fi astfel ridicat!

Argumentaţia matematică din capitolul 2, stabilind forma de care avem nevoie a teoremei lui Gödel, ar putea să pară că are foarte puţin în comun cu cele

mai multe aspecte ale conştiinţei. Într-adevăr, o demonstraţie a faptului că anumite tipuri de înţelegere matematică trebuie să implice ceva dincolo de calcul, nu pare să aibă prea mare importanţă pentru ceea ce este implicat în percepţia culorii roşu, de exemplu şi nici nu pare să existe vreun rol manifest pentru matematică în cele mai multe aspecte ale conştiinţei. De exemplu, atunci când visează, nici chiar matematicienii nu se gândesc, în mod normal, la matematică! Câinii se pare că visează şi se presupune că sunt conştienţi, într-o anumită măsură, când visează; şi eu m-aş gândi că ei sunt conştienţi şi în alte ocazii, dar nu fac matematică. Fără îndoială, contemplarea matematică este foarte departe de a fi singura activitate a animalului care ar cere conştiinţă! Este o activitate eminamente umană, foarte specializată (de asemenea, unii cinici ar putea spune chiar că este o activitate restrânsă la câţiva oameni mai neobişnuiţi). Fenomenul conştiinţei, pe de o parte, este universal, fiind de asemenea prezent în multe activităţi mentale umane precum şi non-umane şi, cu siguranţă, la oamenii care nu sunt matematicieni, ca şi la matematicieni atunci când nu fac matematică (ceea ce se întâmplă în cea mai mare parte a timpului).

Gândirea matematică este o zonă foarte mică din activitatea conştientă pe care şi-o permite o minoritate de fiinţe conştiente pentru o fracţiune din viaţa lor conştientă.

De ce am ales să adresez întrebarea asupra conştiinţei aici, într-un context matematic? Motivul este că doar în interiorul matematicii putem spera să găsim o demonstraţie cât de cât riguroasă a faptului că cel puţin *ceva* din activitatea conştientă *trebuie* să fie necalculabil. Calculul în sine, prin natura sa, este, în fond, matematic. Nu putem spera să găsim o 'demonstraţie' că ceva nu este calcul decât întorcându-ne la matematică. Voi încerca să conving cititorul că tot ceea ce facem cu creierul nostru, sau cu mintea noastră atunci când înţelegem *matematica*, este diferit de tot ceea ce putem obţine de la un calculator; astfel cititorul va fi mai dispus să accepte un rol important al activităţii non-computaţionale, în general, în gândirea conştientă.

Fără îndoială, aşa cum mulţi ar susţine, este evident că senzaţia de 'roşu' nu poate fi evocată nici-cum prin execuţia unor calcule. Dar de ce ne preocupăm atât de serios de existenţa unei demonstraţii matematice inutile când, în mod evident, experienţa subiectivă nu are nimic cu calculul? Un răspuns ar fi că acest *evident* (pentru care eu am o mare simpatie) se referă numai la aspectul pasiv al

conştiinţei. Ca în exemplul Camerei chinezeşti a lui Searle, el poate fi privit ca argument împotriva punctului de vedere $\mathcal{A}$, dar nu-l distinge pe $\mathcal{C}$ de $\mathcal{B}$.

Mai mult, voi ataca modelul de calcul funcţionalist (adică punctul de vedere $\mathcal{A}$) pe teren propriu, ca să spun aşa; ideea funcţionaliştilor este că orice calitate (*qualia*) poate fi produsă, pur şi simplu, prin executarea unor calcule potrivite, fără să conteze cât de improbabil apare un astfel de tablou (la prima vedere). Pentru că ei spun: ce altceva ar fi cu adevărat util să facem cu creierul nostru, dacă nu să executăm calcule de un anumit tip? Ce este creierul dacă nu un anumit tip de sistem comandat de un calculator — de altfel, foarte sofisticat? Ei vor susţine că orice senzaţie de *conştienţă* produsă în vreun fel de activitatea creierului este rezultatul acţiunii calculului. Adeseori adepţii lui $\mathcal{A}$ susţin că, dacă cineva refuză să accepte modelul computaţional pentru *toate* activităţile mentale, inclusiv conştiinţa, acela de fapt face apel la *misticism* (acest lucru ar însemna că singura alternativă la $\mathcal{A}$ este $\mathcal{D}$!).

Intenţia mea este să aduc, în partea a doua a cărţii, unele sugestii parţiale asupra a ce altceva poate face un creier descris ştiinţific. Nu voi nega faptul că unele părţi 'constructive' din argumentaţia mea sunt speculative. În plus, cred că un anumit tip de acţiune non-computaţională este de mare interes şi că pentru a demonstra acest lucru mă voi întoarce la gândirea matematică.

1.19. Ce legătură are teorema lui Gödel cu comportamentul de zi cu zi ?

Să presupunem că acceptăm existenţa unui ingredient nealgoritmic care intervine în judecăţile noastre matematice conştiente, pe baza cărora ajungem la o decizie matematică conştientă. Ce importanţă are acest lucru pentru problema limitelor activităţii unui robot, dacă, aşa cum am menţionat mai sus, se pare că această problemă este legată de bunul-simţ şi nu de comportarea sofisticată a unor experţi antrenaţi?

La prima vedere, concluziile mele pot să pară *opuse* celor găsite pentru limitele inteligenţei artificiale — cel puţin a celor actuale. Deoarece numai aparent eu susţin că o comportarea non-computaţională se poate găsi în zonele sofisticate ale înţelegerii matematice, mai degrabă decât în comportamentul de zi cu zi. Dar

eu nu asta susţin. Eu susţin că *înţelegerea* implică un acelaşi tip de proces nealgoritmic, indiferent dacă se referă la o percepţie matematică pură, să zicem a infinitudinii mulţimii numerelor naturale, sau, pur şi simplu, în observaţia că un obiect de formă dreptunghiulară poate fi folosit pentru a propti fereastra deschisă, sau la înţelegerea modului în care un animal poate fi priponit sau eliberat prin câteva mişcări selectate ale unei bucăţi de frânghie, sau în înţelegerea cuvintelor 'fericire', 'luptă' sau 'mâine', sau în înţelegerea faptului că, atunci când piciorul stâng al lui Abraham Lincoln se afla la Washington, piciorul său drept se afla, aproape sigur, tot la Washington — exemplu care, în mod neaşteptat, încurcă complet un sistem **IA** actual! [26] Acest proces nealgoritmic, ori unde s-ar afla, ne permite să devenim conştienţi de ceva. Această conştienţă ne permite să vizualizăm mişcarea geometrică a unei bucăţi de lemn, sau proprietăţile topologice ale unei bucăţi de sfoară, sau 'conectivitatea lui Abraham Lincoln'. Ne permite de asemenea să avem un acces direct la experienţa unei alte persoane, astfel încât să se poată *şti* ce anume înseamnă cuvinte ca *fericire*, *luptă* sau *mâine*, chiar dacă explicaţiile n-au fost prea adecvate. 'Semnificaţiile' cuvintelor pot trece de la o persoană la alta, nu fiindcă se dau explicaţii adecvate, ci fiindcă cealaltă persoană a avut unele percepţii — sau 'conştientizări' — despre ceea ce pot însemna acele cuvinte, astfel încât chiar explicaţii nepotrivite pot fi suficiente pentru a 'prinde' semnificaţia corectă. Faptul că cele două persoane posedă o conştienţă comună, permite să aibă loc comunicarea între ele. Acest lucru pune robotul comandat de calculator într-un dezavantaj sever (într-adevăr, chiar *semnificaţia* conceptului de 'semnificaţie' a unui cuvânt este ceva despre care avem un anumit tip de înţelegere directă şi e greu să ne imaginăm cum poate da cineva vreo descriere adecvată a *acestui* concept robotului nostru insensibil). Semnificaţiile pot fi comunicate de la o persoană la alta deoarece fiecare persoană are o experienţă interioară sau trăiri de acelaşi fel, despre aceleaşi lucruri. Ne-am putea imagina că *experienţele* sunt lucruri care constituie un anumit tip de memorie care înregistrează ce anume s-a întâmplat şi că robotul nostru ar putea fi cu uşurinţă echipat cu aşa ceva. Dar sunt de părere că lucrurile nu stau aşa şi că este esenţial ca obiectul în chestiune, o fiinţă umană sau un robot, să fie efectiv conştient de experienţă.

De ce spun că această *conştienţă*, ori ce ar fi ea, trebuie să fie non-computaţională, astfel încât nici un robot comandat de un calculator, bazat pur şi simplu pe logica standard a maşinii Turing (sau altceva echivalent) — cu algoritmi

top-down sau *bottom-up* — n-ar putea-o dobândi sau *simula*? Acesta este punctul în care raţionamentul de tip Gödel joacă un rol esenţial. Este greu de zis astăzi ceva despre 'a fi conştient' de, să zicem, culoarea roşie; dar *există* ceva definit ce putem spune despre 'a fi conştient' de infinitatea numerelor naturale. Este *conştienţa* care permite unui copil să *ştie* ce se înţelege prin numerele *unu*, *doi*, *trei*, *patru* etc. şi ce se înţelege prin faptul că acest şir continuă indefinit, atunci când i s-au dat doar câteva explicaţii absurd de limitate şi aparent irelevante în termenii câtorva mere sau pere. Conceptul *trei* poate fi abstractizat de copil din astfel de exemple limitate; şi mai mult, copilul poate să prindă faptul că acest concept este unul dintr-un şir fără sfârşit de concepte asemănătoare (*patru*, *cinci*, *şase*, etc.) Într-un anumit sens platonic, copilul deja *ştie* ce sunt numerele naturale.

Acest lucru pare puţin cam mistic, dar nu este de fapt. Este vital pentru discuţia care urmează, să distingem acest tip de cunoştinţe platonice, de misticism. Conceptele pe care le *ştim*, în sens platonic, sunt lucrurile 'evidente' pentru noi — lucrurile care pot fi reduse la perceperea pe baza bunului simţ — deşi poate nu suntem în stare să caracterizăm complet aceste concepte în termenii unor reguli de calcul. Într-adevăr, aşa cum vom vedea în legătură cu raţionamentul lui Gödel, în cele ce urmează, nu există nici un mod de a caracteriza complet proprietăţile numerelor naturale, în termenii unor astfel de reguli. Totuşi, cum sunt descrise numerele cu ajutorul merelor sau a bananelor încât să permită copilului să ştie ce sunt 'trei zile', adică să ştie că acelaşi concept abstract de *trei* este prezent şi în expresia 'trei pere'? Evident, această capacitate nu apare dintr-o dată şi copilul ar putea să se înşele la început, dar asta nu este esenţial. Esenţial este faptul că acest tip de înţelegere este totuşi posibilă. Conceptul abstract de *trei* şi faptul că acesta este unul dintr-un şir infinit de concepte corespunzătore — chiar numerele naturale însele — este ceva ce poate fi înţeles, dar, pretind eu, doar utilizând conştienţa.

Eu pretind că noi nu folosim reguli de calcul atunci când vizualizăm o bucată de lemn, un capăt de sfoară, sau pe Abraham Lincoln. De fapt, există simulări pe calculator foarte bune ale mişcării unui corp rigid, ca o bucată de lemn, de exemplu. Simulările acestor mişcări pot fi făcute atât de precis şi demn de încredere, încât ele pot fi cu mult mai fidele decât ceea ce se poate atinge prin vizualizare umană directă. La fel, mişcarea unei bucăţi de sfoară sau a unui şiret poate fi simulată prin calcul, deşi, surprinzător, acest lucru este mult mai dificil decât simularea mişcării unui corp solid (o explicaţie parţială este faptul că 'o

sfoară matematică' necesită un număr infinit de parametri pentru specificarea poziţiei sale în vreme ce corpul rigid necesită doar şase). Există un algoritm de calcul pentru a decide dacă sfoara este înnodată, dar este complet diferit de cel care descrie mişcarea corpului rigid (şi nu este foarte eficient din punctul de vedere al calculului). Orice simulare a lui Abraham Lincoln ar fi mult mai dificilă. Însă esenţial pentru mine nu este dacă 'vizualizarea' umană este 'mai bună' sau 'mai rea' decât a unui calculator pentru aceste lucruri variate.

O observaţie importantă: mi se pare că vizualizarea implică un element de apreciere a ceea ce este de văzut; adică implică *înţelegere*. Pentru a ilustra acel lucru pe care îl am în minte, să considerăm un fapt elementar din aritmetică şi anume că, pentru oricare două numere naturale (adică numere întregi nenegative: 0, 1, 2, 3, 4,...) a şi b, avem proprietatea:

$$a \times b = b \times a$$

Trebuie să fie clar că aceasta nu este o egalitate trivială pentru că semnificaţia celor două părţi ale egalităţii este diferită. La stânga, $a \times b$ se referă la o colecţie de a grupe a câte b obiecte; în vreme ce $b \times a$ din dreapta se referă la b grupe de câte a obiecte. În cazul particular în care $a=3$ şi $b=5$ avem pentru $a \times b$ o reprezentare de forma:

(●●●●●)(●●●●●)(●●●●●) ,

în vreme de pentru $b \times a$ avem

(●●●)(●●●)(●●●)(●●●)(●●●)

Faptul că avem acelaşi număr de puncte în fiecare caz demonstrează că $3 \times 5 = 5 \times 3$.

Putem vedea că acest lucru este adevărat vizualizând mulţimea

●●●●●
●●●●●
●●●●●

Dacă numărăm punctele pe linii vedem că avem trei linii, fiecare cu câte cinci puncte, ceea ce se exprimă cantitativ ca 3×5. Dar dacă numărăm punctele pe coloane, găsim că avem cinci coloane cu câte trei elemente, exprimat cantitativ prin 5×3 Faptul că cele două cantităţi sunt egale se vede dintr-o dată deoarece este vorba de aceeaşi mulţime de puncte aşezate într-un dreptunghi, în fiecare caz (eventual, am putea prefera să rotim dreptunghiul cu un unghi drept şi cu ochii minţii să vedem că mulţimea reprezentând 5×3 are acelaşi număr de elemente ca

cea care reprezintă pe 3×5).

Ceea ce este important în acest act de vizualizare este că el ne dă direct mult mai mult decât faptul particular că $3\times5=5\times3$. Cum valorile particulare ($a=3$ şi $b=5$) sunt oarecare în această procedură, putem lua, să zicem $a=79797000222$ şi $b=50000123555$ şi putem susţine cu hotărâre că

$$79797000222\times5000012355 = 50000123555\times79797000222$$

în ciuda faptului că nu avem nici o şansă să vizualizăm un dreptunghi atât de mare (şi nici un calculator actual nu poate să-i numere elementele). Putem să tragem concluzia că egalitatea de mai sus trebuie să aibă loc — sau că egalitatea mai generală[13] $a\times b=b\times a$ trebuie să aibă loc — în esenţă prin aceiaşi metodă de vizualizare pe care am folosit-o pentru $3\times5=5\times3$. În principal trebuie să 'estompăm' în mintea noastră numărul precis al liniilor şi coloanelor care se folosesc şi egalitatea devine evidentă.

Nu vreau să susţin că toate relaţiile matematice pot fi percepute direct ca 'evidente' dacă sunt vizualizate într-un mod potrivit- sau că pur şi simplu ele pot fi totdeauna percepute într-un mod care să fie imediat clar pentru intuiţia noastră. Departe de mine acest gând. Anumite relaţii matematice necesită lungi lanţuri de raţionament înainte de a fi percepute ca certitudine. Dar obiectul demonstraţiei matematice este, de fapt, să furnizeze astfel de lanţuri de raţionamente în care fiecare *pas* este cu adevărat ceva care poate fi perceput ca evident. În consecinţă, punctul final al raţionamentului este ceva ce trebuie acceptat ca *adevărat*, chiar dacă în sine el nu este deloc evident.

Ne putem imagina că am putea să facem o listă cu toţi paşii 'evidenţi' ai raţionamentului, odată pentru totdeauna, astfel încât totul să se reducă la calcul — adică la manipularea mecanică a acestor paşi evidenţi. Raţionamentul lui Gödel arată că acest lucru nu este posibil (vezi paragraful 2.5). Nu există nici un mijloc prin care să se poată elimina nevoia de a 'înţelege' *noi* lucruri 'evidente'. Astfel, înţelegerea matematică nu poate fi redusă la un calcul orb.

---

[13] Trebuie să remarcăm că această egalitate *nu* are loc pentru diferite feluri de numere 'stranii' care intervin în matematică, ca de exemplu, numerele ordinale la care ne referim în paragraful 2.10, după Q19. Dar egalitatea are loc întotdeauna pentru numerele *naturale*, cele cu care ne-am ocupat până aici.

1.20. Vizualizarea mentală şi realitatea virtuală

Intuiţiile matematice schiţate în paragraful 1.19 au avut un caracter mai mult geometric. Există însă şi multe alte intuiţii care pot fi utilizate în matematică — intuiţii care nu trebuie să fie neapărat geometrice. Totuşi, se constată că intuiţiile geometrice au adeseori o valoare deosebită în înţelegerea matematică. Astfel, ar fi instructiv să ne întrebăm ce fel de activitate fizică are loc de fapt în creierul nostru când vizualizăm ceva într-un mod geometric.

Nu există nici o justificare logică pentru ca vizualizarea să fie ea însăşi 'o oglindire geometrică' a lucrurilor care sunt de vizualizat. Aşa cum vom vedea, acest proces mental ar putea să fie ceva complet diferit.

Este util să facem o comparaţie cu ceea ce se susţine că are relevanţă pentru 'vizualizare'. Conform procedurilor din realitatea virtuală [27], simularea pe calculator este făcută din structuri care nu există — analog proiectului (arhitectural) unei clădiri — şi această simulare este propusă ochiului unui subiect uman, care pare să perceapă structura ca 'reală'. Prin mişcări ale ochiului sau capului, sau chiar a picioarelor care s-ar plimba în jur, subiectul va putea vedea structura din diferite unghiuri, aşa cum s-ar întâmpla dacă structura ar fi reală (vezi figura 1.8).

Unii cercetători[28] au susţinut că orice s-ar întâmpla în creierul nostru, când vizualizăm conştient, trebuie să fie vorba de ceva foarte asemănător cu calculul din construcţia unei astfel de simulări. Într-adevăr, când este văzută o structură adevărată, 'ochiul minţii' pare că face un model mental care rămâne neschimbat în ciuda mişcărilor capului, ochilor şi corpului (ceea ce ne-ar putea conduce la concluzia că imaginile de pe retină sunt permanent schimbate). Aceste corecţii, legate de mişcările corpului, fac parte, într-o mare măsură, din realitatea virtuală şi au sugerat că ceva foarte asemănător trebuie să se întâmple la construcţia 'modelelor mentale' care constitue de fapt activitatea noastră de vizualizare. Aceste calcule nu trebuie să aibă vreo relaţia geometrică (sau de 'oglindire') cu structura care trebuie modelată.

Suporterii concepţiei $\mathcal{A}$ ar putea să privească vizualizarea conştientă ca fiind rezultatul unei astfel de simulări prin calcul a lumii exterioare, care s-ar petrece în capul nostru. Dimpotrivă, eu susţin că atunci când noi percepem o scenă

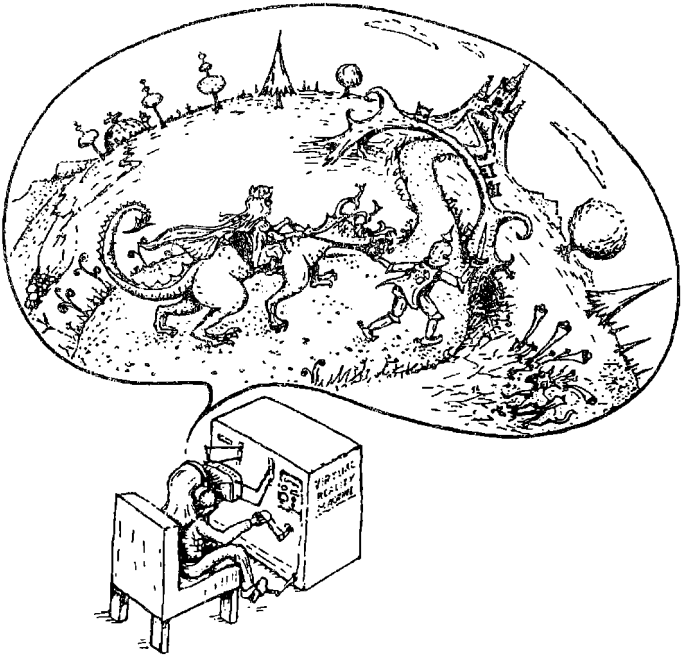


Fig. 1.8. Realitatea virtuală. O simulare tri-dimensională a lumii răspunzând computaţional mişcărilor capului şi corpului.

vizuală, *înţelegerea* implicată este ceva complet diferit de modelarea lumii în termenii unei simulări prin calcul.

S-ar putea susţine că ceva din interiorul creierului nostru lucrează mai mult ca un calculator analogic, în care modelarea lumii exterioare se obţine nu prin calcul numeric, ca în cazul calculatoarelor electronice moderne, ci în termenii unei structuri fizice a cărei comportare fizică s-ar putea traduce ca oglindind comportarea sistemului extern care trebuie modelat.

Dacă am dori să facem un dispozitiv analogic care să modeleze mişcarea unui corp rigid exterior, este foarte clar că există un mod direct de a face acest lucru. Spre exemplu, să avem, intern, un mic obiect fizic de aceeaşi formă (dar de alte dimensiuni) ca obiectul exterior ce trebuie modelat — deşi nu voi sugera în nici un fel că acest model particular are vreo relevanţă pentru ceea ce se întâmplă

de fapt în creier! Mişcarea acestui obiect intern poate fi văzută sub diferite unghiuri, ducând la efecte aparent foarte asemănătoare cu cele obţinute de calculatorul numeric. Un astfel de sistem poate fi folosit, de asemenea, ca parte a unui sistem de *realitate virtuală*, unde în loc să avem un întreg model calculatoriu al structurii în chestiune, am avea un model fizic diferind de *realitatea* care trebuie simulată doar prin mărime.

De fapt simularea analogică nu trebuie să fie atât de directă şi de banală. Poate fi folosit un parametru, ca potenţialul electric, mai degrabă, decât distanţa fizică, sau altă mărime fizică. Trebuie însă să fim siguri că legile fizicii care guvernează structura internă oglindesc foarte precis legile fizicii care guvernează structura externă care trebuie modelată. Nu este necesar ca structura internă să *semene* în vreun fel cu cea externă (sau s-o 'oglindească').

S-ar putea ca dispozitivele analogice să realizeze lucruri inaccesibile calculatoarelor numerice? Aşa cum am susţinut în paragraful 1.8, nu există nici un motiv să credem că, în cadrele actuale ale fizicii, simularea analogică ar putea obţine lucruri pe care simularea numerică nu le-ar putea face. Astfel, dacă argumentele noastre ne sugerează că imaginaţia noastră vizuală implică lucruri nealgoritmice, atunci suntem încurajaţi să privim în afara cadrelor fizicii existente pentru explicarea imaginaţiei noastre vizuale.

1.21. Este imaginaţia matematică nealgoritmică ?

Nimic din cele discutate anterior nu spune explicit că tot ceea ce facem atunci când vizualizăm ceva nu poate fi simulat cu un calculator numeric. S-ar putea ca atunci când utilizăm un anumit tip intern de sistem analogic pentru vizualizare să existe cel puţin posibilitatea de a *simula*, numeric, comportarea dispozitivului analogic.

Vizualizarea la care tocmai m-am referit este legată de ceea ce este *vizual* în cel mai larg sens, adică de imaginea mentală care pare să corespundă semnalelor care ajung la creier de la ochi. Mai general, o imagine mentală nu trebuie deloc să aibă acest caracter *vizual*; în sens literal, ca atunci când cineva înţelege semnificaţia unui cuvânt abstract sau îşi aminteşte o bucată muzicală. Imaginile mentale ale cuiva care este orb din naştere, de exemplu, pot avea cu greu vreo

relaţie directă cu semnalele primite de la ochi. Astfel, 'vizualizarea' la care mă refer, pare să aibă, în general, mult mai mult în comun cu *conştienţa*, decât cu elementele care privesc sistemul vizual. Nu cunosc nici un argument care să aibă vreo legătură directă cu natura puterii noastre de vizualizare, algoritmică sau nealgoritmică, în sensul literal al cuvântului *vizual*. Convingerea mea, că puterea noastră de vizualizare trebuie într-adevăr să fie nealgoritmică, este dedusă din faptul că şi alte forme de conştienţă umană par să aibă un caracter nealgoritmic, demonstrabil.

Deşi este greu să-mi închipui cum s-ar putea da o justificare directă pentru necalculabilitatea specifică vizualizării geometrice, o argumentaţie convingătoare, că anumite forme ale conştiinţei conştiente trebuie să fie nealgoritmice, ar conduce la sugestia că tipul de conştienţă care este implicat în vizualizarea geometrică ar trebui să fie, de asemenea, necalculabil. S-ar părea că nu avem nici un motiv să credem că o linie clară s-ar putea trage între diferitele manifestări ale înţelegerii conştiente, în legătură cu aceste lucruri.

Mai precis, conştienţa, despre care eu susţin că este non-computaţională în mod demonstrabil, este înţelegerea proprietăţilor numerelor naturale 0, 1, 2, 3,... (s-ar putea spune că conceptul nostru de număr natural *este*, într-un anumit sens, o formă de 'vizualizare' *non*-geometrică). Vom vedea în paragraful 2.5, cu ajutorul unei forme cu adevărat accesibile a teoremei lui Gödel, că această înţelegere este ceva ce nu poate fi încapsulat în nici o mulţime finită de reguli — de unde rezultă că nu poate fi simulată cu un calculator (vezi răspunsul la întrebarea **Q16**). Din când în când, auzim că un calculator a fost 'antrenat' să 'înţeleagă' conceptul de număr natural[29]. Totuşi, aşa cum vom vedea, acest lucru nu poate fi adevărat. *Conştienţa* noastră, despre ce anume înseamnă un 'număr', este cea care ne permite să prindem corect conceptul. Atunci când avem format acest concept, în mod corect, putem — cel puţin în principiu — să dăm răspunsurile corecte la o întreagă familie de întrebări despre numere, care ne-ar fi puse, fără să existe formulată vreo mulţime finită de reguli pentru a face aşa ceva. Pe de altă parte, doar cu reguli şi fără conştienţă directă, un robot controlat de un calculator (ca Deep Thought din paragraful 1.15) va fi în mod necesar limitat, într-un mod în care noi nu suntem limitaţi — chiar dacă, dându-i robotului reguli de comportare destul de bune, el poate executa sarcini prodigioase, unele depăşind cu mult posibilităţile umane, în zone înguste, bine definite şi ar putea să ne facă să

credem, pentru o vreme, că el ar avea chiar conştienţă.

Este important să subliniem că, atunci când se realizează o simulare efectivă cu un calculator numeric (sau analogic) a unui sistem exterior, aproape totdeauna acest lucru se face pe baza înţelegerii profunde a unor idei matematice. Să considerăm simularea mişcării geometrice a unui corp rigid. Calculele necesare depind, în mod esenţial de intuiţiile unor mari gânditori ai secolului şaptesprezece, ca acei matematicieni francezi care au fost Descartes, Fermat şi Desargues, cei care au introdus ideile geometriei analitice şi proiective. Ce se poate spune despre simularea mişcării unei bucăţi de sfoară? Ideile geometrice necesare pentru a înţelege restricţiile comportării unei bucăţi de sfoară — adică faptul că ea se poate 'înnoda' — sunt foarte sofisticate; şi ele au apărut destul de recent, multe dintre progresele fundamentale având loc în acest secol. Deşi, în practică, nu este prea dificil să decidem, manipulând cu o singură mână o sfoară şi aplicând o înţelegere de bun simţ, dacă o buclă a sforii este sau nu înnodată, totuşi, algoritmul de calcul pentru realizarea aceluiaşi lucru, este surprinzător de dificil, sofisticat şi încă insuficient.

Astfel, simularea numerică efectivă a unor astfel de situaţii a fost bazată pe algoritmi *top-down* şi a depins considerabil de înţelegerea şi intuiţia umană. Este foarte puţin probabil ca ceva similar să se petreacă în creierul nostru atunci când este angajat într-un act de vizualizare. O variantă mult mai plauzibilă ar fi ca ingrediente *bottom-up* importante să fie implicate, astfel încât 'imaginea vizuală' să apară numai după ce a avut loc o considerabilă activitate de 'învăţare din experienţă' Nu ştiu dacă există abordări semnificative, *bottom-up* (de exemplu cu reţele neuronale), ale problemelor de acest tip. Cred totuşi că o abordare care ar fi în întregime bazată pe o organizare *bottom-up* ar da rezultate proaste. Este greu de crezut că ar putea fi dată o simulare bună o mişcării geometrice a corpului rigid sau a restricţiilor topologice ale mişcării unei sfori — adică a înnodării — fără să fie implicată o înţelegere reală, a ceea ce este de făcut, de fapt.

Ce fel de proces fizic ar putea fi acela care să fie responsabil de conştienţa noastră — o conştienţă care pare că este necesară pentru orice înţelegere reală? Ar putea fi ceva deasupra simulării computaţionale, aşa cum cere punctul de vedere $\mathcal{C}$? Este acest presupus proces fizic ceva care, prin el însuşi, să fie accesibil înţelegerii noastre, cel puţin în principiu? Eu cred că da şi că punctul de vedere $\mathcal{C}$ este o posibilitate ştiinţifică autentică, deşi trebuie să fim pregătiţi pentru

eventualitatea ca metodele şi criteriile ştiinţifice să sufere importante modificări. Noi va trebui să fim pregătiţi să examinăm variante neaşteptate, plasate în zone ale înţelegerii pure care pot, la prima vedere să pară complet irelevante. În discuţiile care vor urma, îl rog pe cititor să fie deschis la argumente şi, în acelaşi timp, să fie foarte atent la raţionament şi la rigoarea ştiinţifică, chiar atunci când i se va părea că demonstraţia vine în conflict cu ceea ce se consideră evident, din punctul de vedere a simţului comun. Fiţi gata să vă gândiţi puţin la argumentele pe care mă străduiesc să le prezint (cât de clar voi putea). Astfel încurajaţi să ne aventurăm înainte.

În partea întâi, voi lăsa de-o parte problemele de fizică şi biologie care ar putea sta la baza non-computabilităţii cerute de punctul de vedere $\mathcal{C}$. Aceste chestiuni vor fi tratate în partea a doua a cărţii. Dar de ce sunt necesare astfel de acţiuni non-computaţionale? Necesitatea lor se datorează, spun eu, faptului că acţiunea noastră este nealgoritmică atunci când înţelegem ceva în mod conştient. Trebuie să justific această pretenţie şi, de aceea, trebuie să ne întoarcem către matematica noastră.

Note şi referinţe bibliografice

1. Vezi, în particular, Good (1965), Minsky (1986), Moravec (1988).
2. Moravec (1988) îşi bazează argumentaţia pentru această estimare, pe proporţia proceselor din cortex care consideră că au fost deja modelate cu succes (mai ales cele de pe retină), împreună cu o apreciere a ritmului de dezvoltare (în viitor) a tehnologiei informatice. La începutul lui 1994 (data finalizării cărţii de fa-ţă) el încă susţinea aceste estimări; cf. Moravec (1994).
3. Aceste patru puncte de vedere au fost descrise explicit, spre exemplu, în Johnson-Laird (1987), p.252 (cu menţiunea că aceştia numesc 'teza Church-Turing' ceea ce aici în carte, în paragraful 1.6, autorul numeşte 'teza lui Turing', spre deosebire de 'teza lui Church').
4. Spre exemplu, D. Dennett, D. Hofstadter, M. Minsky, H. Moravec, H. Simon; pentru o discuţie asupra acestor termeni se poate consulta Searle (1980), Lockwood (1989).
5. Vezi Moravec (1988).
6. Turing (1950); vezi **MN**, paginile 15-20.

7. Vezi Searle (1980), (1992).
8. Afirmaţia este încurcată de faptul că fizica zilelor noastre este dependentă mai degrabă de folosirea *continuului*, decât de acţiunile discrete (numerice), chiar dacă sensul *computabilităţii* în acest context este deschis diferitelor interpretări. Pentru diferite discuţii relevante se pot consulta Pour-el (1974), Smith şi Stephenson (1975), Pour-el şi Richards (1979), (1981), (1982), (1989), Blum, Shub şi Smale (1989), Rubel (1988), (1989). Subiectul va fi reluat în paragraful 1.8.
9. Am auzit această frază drăguţă la Radio BBC 4, în emisiunea 'Gândul zilei' (Thought of the Day).
10. Problematica Inteligenţei Artificiale a apărut efectiv încă în anii '50, prin folosirea comparativă a procedurilor elementare *top-down* (spre exemplu, Grey Walter 1953). Schema *perceptron* ca instrument de recunoaştere a formelor, elaborată de Frank Rosenblatt (1962), în 1959, a fost primul aparat *conecţionist* de succes (reţea neuronală), aceasta stimulând un mare interes pentru scheme *bottom-up*. Desigur, încă din 1969, Marvin Minsky şi Seymour Papert (cf. Minsky şi Papert 1972) au subliniat limitările esenţiale ale acestei organizări *bottom-up*. Acestea au fost mai târziu depăşite de Hopfield (1982), iar aparatele de tip reţea neuronală sunt acum subiectul unor considerabile cercetări din toată lumea (Vezi, spre exemplu, Beks şi Hamker 1992 şi Gernoth şi alţii 1993, pentru unele aplicaţii în fizica energiilor înalte). Ca repere importante în cercetările **IA** din zona algoritmilor *top-down* pot fi considerate lucrările lui John McCarthy (1979) şi Alan Newell and Herbert Simon (1976). Vezi Freedman (1994) — o prezentare dramatică a întregii istorii a **IA**. Pentru alte discuţii recente, privind metodele şi rezultatele **IA**, se pot consulta Grossberg (1987), Baars (1988); pentru contestările domeniului se poate vedea Dreyfus (1972); iar pentru punctele de vedere recente ale unui pionier al **IA**, Gelernter (1994); cf. diferitelor articole din Broadbent (1993) şi Khalfa (1994).
11. Pentru o prezentare a λ-calculului vezi Church (1941) şi Kleene (1952).
12. Dintre diversele publicaţii privind acest domeniu se pot consulta, ca repere bibliografice, Pour-el (1974), Smith and Stephenson (1975), Pour-el and Richards (1989), Blumb, Shub, and Smale (1989). În particular, problema activităţii creierului în relaţie cu aceste chestiuni a fost tratată de Rubel (1985).
13. In cazul problemei de acoperire , ceea ce a demonstrat de fapt Robert Berger a fost că pentru acoperirile Wang problema nu are soluţie algoritmică generală. Acoperirile Wang (după numele logicianului Hao Wang) constau din pătrate de acoperire cu muchii colorate, la care culoarea acestor muchii trebuie să coincidă la îmbinări (rotaţia şi oglindirea nefiind deci permise). Nu este greu de dedus că, pentru orice

mulţime de acoperiri Wang, setul corespunzător de structuri de acoperire (*polyominoes*) acoperă planul dacă şi numai dacă mulţimea pătratelor Wang îl acoperă. În concluzie, insolvabilitatea prin calcul a problemei de acoperire rezultă direct şi imediat din cea a acoperirilor Wang. Este important să menţionăm că dacă o mulţime dată de *polyominoes* nu poate acoperi planul, atunci acest fapt poate fi determinat computaţional (în stilul opririi maşinii Turing, sau ca atunci când un set de ecuaţii diofantiene admite o soluţie), aşa cum încercarea de a acoperi o regiune $n \times n$ pentru valori tot mai mari ale lui n, va eşua de la o anumită valoare a acestuia. Există situaţii în care acoperirea planului nu poate fi determinată computaţional.

14. Vezi Freedman (1994) pentru o imagine foarte optimistă a aspiraţiilor iniţiale din **IA**.
15. Trebuie să mulţumesc multor specialişti şi, în particular, lui Lee Loevinger pentru că m-a familiarizat cu aceste chestiuni. Vezi şi Hodgson (1991) pentru o discuţie remarcabilă asupra relevanţei fizicii moderne şi a calculului asupra întrebării: cum ne comportăm.
16. Vezi, spre exemplu, Smithers (1990).
17. De exemplu, Sloman (1992) m-a admonestat pentru că am pus atât de multe consideraţii în sarcina termenului prost-definit de *consciousness* în **MN**, în timp ce el se referă mereu la şi mai puţin bine definitul termen *mind*!
18. Searle (1980), (1992).
19. Vezi la pagina 372 volumul editat de Hofstadter şi Dennett (1981) articolul lui Searle (1980). Nu mi-e prea clar, desigur, de ce Searle susţine acum punctul de vedere $\mathcal{B}$ şi nu pe $\mathcal{C}$.
20. Vezi Hofstadter (1981) pentru o prezentare amuzantă şi antrenantă a unei astfel de sugestii; vezi şi p. 30 din **MN**.
21. Pentru o introducere accesibilă asupra noţiunii de *complexitate algoritmică* vezi Chaitin (1975).
22. Vezi Hsu şi alţii (1990).
23. Vezi Freedman (1994).
24. Vezi, spre exemplu, Moravec (1994).
25. Vezi Putnam (1960), Smart (1961), Benacerraf (1967), Good (1967, 1969), Lewis (1969, 1989), Hofstadter (1981), Bowie (1982), în legătură cu argumentele lui Lucas; vezi deasemeni Lucas (1970). Versiunea mea proprie, prezentată succint în **MN** paginile 450-452, a fost atacată în diferite recenzii; în particular Sloman (1992) şi numeroşi comentatori în *Behavioral and Brain Sciences*: Boolos (1990), Butterfield (1990), Chalmers (1990), Davis (1990), (1993), Dennett (1990), Doyle (1990), Glymour and Kelly

(1990), Hodgkin and Houston (1990), Kentridge (1990), MacLennan (1990), McDermott (1990), Manaster–Ramer şi alţii (1990), Mortensen (1990), Perlis (1990), Roskies (1990), Tsotsos (1990), Wilensky (1990); vezi şi replicile mele: Penrose (1990, 1993d) şi, deasemeni, Guccione (1993); vezi şi Dodd (1991), Penrose (1991b).

26. Dintr–un program TV britanic – probabil „The Dream Machine" din decembrie 1991, a 4–a parte a seriei BBC „The Thinking Machine". Vezi şi Freedman (1994) pentru o discuţie a progreselor recente din **IA** 'Inţelegere', în particular referitor la senzaţionalul proiect *Cyc* al lui Douglas Lenat.
27. Pentru o introducere clară şi simplificatoare, vezi Woolley (1992).
28. Spre exemplu, o astfel de sugestie a fost făcută de Richard Dawkins în lecţiile sale de Crăciun la BBC 1992.
29. Vezi, spre exemplu, interpretarea lui Freedman (1994) privind lucrarea lui Lenat şi alţii în această direcţie.

2. EXEMPLUL LUI GÖDEL

2.1. Teorema lui Gödel şi maşinile Turing

În matematică, procesele noastre de gândire au forma lor cea mai pură. Dacă gândirea înseamnă doar efectuarea unui calcul de un anumit tip, atunci ar trebui să observăm cel mai clar acest lucru în gândirea noastră matematică. Totuşi, în mod remarcabil, situaţia pare să fie exact invers. Tocmai în matematică găsim cea mai clară dovadă a faptului că există ceva care nu se reduce la calcul. Acest lucru pare paradoxal dar va fi de primă importanţă în argumentarea care urmează.

Înainte de a începe, permiteţi-mi să încurajez cititorul să nu se sperie de matematica pe care o voi folosi în următoarele câteva secţiuni (paragrafele 2.2 2.5), în ciuda faptului că vom ajunge să facem, nici mai mult nici mai puţin, decât aprecieri asupra implicaţiilor celei mai importante teoreme din logica matematică din toate timpurile, faimoasa teoremă a lui Kurt Gödel. Voi prezenta doar o formă extrem de simplificată a acestei teoreme, îndreptându-mi apoi atenţia asuprea ideilor ceva mai târzii ale lui Alan Turing. Nu voi utiliza decât formalismul matematic al aritmeticii celei mai simple. Admit că raţionamentul pe care-l fac va fi încurcat pe alocuri, dar va fi doar încurcat nu 'dificil' şi nu va cere nici un fel de cunoştinţe matematice prealabile. Citiţi argumentaţia cât de încet doriţi şi nu vă jenaţi să recitiţi de câte ori doriţi. Mai târziu (în paragrafele 2.6 - 2.10), vom analiza câteva dintre aspectele care stau la baza teoremei lui Gödel, dar cititorul care nu este interesat de aceste chestiuni nu trebuie să-şi bată capul cu această parte a cărţii.

Ce este, de fapt, teorema lui Gödel ? În 1930 tânărul şi strălucitul matematician Kurt Gödel a şocat, la o întâlnire la Königsberg, un grup de matematicieni şi logicieni dintre cei mai importanţi din lume, cu ceea ce urma să fie celebra sa teoremă. A fost rapid acceptată ca fiind o contribuţie esenţială la fundamentele matematicii – poate chiar cea mai importantă – dar eu voi spune mai mult şi anume că, stabilind această teoremă, Gödel a realizat un pas important în filosofia minţii.

Printre lucrurile pe care Gödel le-a stabilit este acela că nici un sistem formal respectând regulile matematice ale demonstraţiei corecte nu ar putea , nici chiar în principiu, să demonstreze toate propoziţiile adevărate din aritmetica obişnuită. Acest lucru este remarcabil în sine. Dar rezultatul său arată mai mult decât atât, el stabileşte că înţelegerea umană şi discernământul nu pot fi reduse la nici o mulţime de reguli de calcul.

Ceea ce rezultă din teorema lui Gödel este că nici un sistem de reguli nu poate fi vreodată suficient pentru a demonstra nici chiar propoziţiile aritmeticii al căror adevăr este accesibil în principiu intuiţiei umane şi discernământului – că deci intuiţia umană şi discernământul nu pot fi reduse la nici un set de reguli. Intenţia mea este să-l conving pe cititor că într-adevăr teorema lui Gödel arată acest lucru şi furnizează fundamentul argumentaţiei mele că trebuie sa existe mai mult în gândirea umană decât s-ar putea vreodată realiza de către un calculator, în sensul în care înţelegem astăzi termenul 'calculator'.

Nu este necesar să dăm acum definiţia unui sistem formal pentru partea principală a argumentaţiei (vezi totuşi paragraful 2.7). În loc de aceasta, vom folosi contribuţiile lui Turing de prin 1936 (şi ale altora; în primul rând ale lui Church şi Post) care au fundamentat tipul de proces pe care noi îl numim astăzi *calcul* sau 'algoritm' Există o echivalenţă efectivă între astfel de procese şi ceea ce poate fi realizat de către un sistem matematic formal. Deci nu va fi important să ştim ce este de fapt un sistem formal cu condiţia să avem o idee suficient de clară despre ce se înţelege printr-un calcul sau un algoritm. Nici chiar pentru acestea nu sunt necesare definiţii riguroase.

Acei cititori care sunt familiarizaţi cu cartea mea precedentă (MN cap.2), ştiu că un algoritm este ceea ce poate fi calculat de o *maşină Turing*, la care ne

putem gândi ca la un calculator ideal. O maşină Turing execută activităţile printr-o procedură pas cu pas, fiecare pas fiind complet specificat de un semn de pe o 'bandă' pe care maşina o examinează în fiecare moment şi de 'starea internă' (definită discret) a maşinii. Diferitele stări interne permise sunt în număr finit, şi numărul total al semnelor de pe bandă este de asemenea finit — deşi banda în sine poate fi infinit de lungă.

Maşina porneşte dintr-o stare particulară, să zicem dintr-una etichetată „*0*", iar instrucţiunile sunt introduse de pe bandă, să zicem sub forma unui număr binar (şiruri de 0 şi 1). Maşina începe să citească aceste instrucţiuni, să mişte banda (sau echivalent, să se mişte chiar ea de-lungul benzii) într-un mod definit de procedura pas cu pas, mod care este determinat la fiecare pas de starea internă şi de cifra particulară care se întâmplă să fie examinată la intrare. Maşina şterge semne, sau face altele, tot în conformitate cu procedura dată. Şi continuă în acest mod până ajunge la o instrucţiune : „STOP" — şi în acest punct, şi numai în acest punct, este afişat rezultatul calculului pe bandă şi activitatea maşinii este terminată. Maşina este acum gata pentru un nou calcul.

Anumite maşini Turing sunt numite *universale*, acestea fiind maşini Turing care au capacitatea de a imita orice altă maşină Turing. Astfel o singură maşină Turing universală are capacitatea să efectueze orice calcul (sau algoritm) care poate fi specificat.

Deşi construcţia internă a calculatoarelor moderne este foarte diferită de aceasta (şi spaţiul de memorie, deşi foarte mare, nu este infinit ca al benzii ideale a maşinii Turing), toate calculatoarele universale sunt, de fapt, maşini Turing universale.

2.2. Proceduri de calcul

Ne vom ocupa în continuare de *proceduri de calcul*. Înţeleg prin procedură de calcul (sau algoritm) acţiunea unei maşini Turing, adică operaţiile unui calculator executate în conformitate cu un program. Trebuie să observăm că procedurile de calcul nu sunt doar un şir de operaţii aritmetice, ca adunarea şi înmulţirea numerelor, ci implică şi alte lucruri. Astfel, operaţiile logice bine definite pot fi şi ele parte a procedurii de calcul. Ca exemplu de calcul putem

considera următoarea problemă:

(A) Să se găsească un număr care nu este suma a trei pătrate perfecte.

Prin număr înţelegem aici un număr natural adică unul dintre numerele:

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,

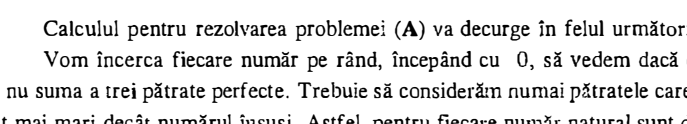
Un pătrat perfect este un număr care se obţine ca produs al unui număr natural cu sine însuşi, de exemplu:

0, 1, 4, 9, 16, 25, 36,

acestea fiind respectiv:

$0\times0=0^2$, $1\times1=1^2$, $2\times2=2^2$, $3\times3=3^2$, $4\times4=4^2$, $5\times5=5^2$, $6\times6=6^2$,

Aceste numere se numesc pătrate pentru că pot fi reprezentate ca



Calculul pentru rezolvarea problemei (**A**) va decurge în felul următor:

Vom încerca fiecare număr pe rând, începând cu 0, să vedem dacă este sau nu suma a trei pătrate perfecte. Trebuie să considerăm numai pătratele care nu sunt mai mari decât numărul însuşi. Astfel, pentru fiecare număr natural sunt doar un număr finit de numere care trebuiesc încercate.

De îndată ce găsim trei pătrate care însumate dau numărul nostru, o etapă a calculului este încheiată şi putem trece la numărul următor. Vom încerca, pentru următorul număr, să găsim din nou trei pătrate perfecte (toate cel mult egale cu numărul dat) a căror sumă să fie chiar numărul dat. Calculul nostru se va încheia numai atunci când găsim un număr natural pentru care un astfel de triplet de pătrate nu există.

Pentru a vedea cum se lucrează să începem cu 0. Acesta este $0^2+0^2+0^2$ şi este cu adevărat suma a trei pătrate. Mai departe să încercăm cu 1, care, deşi nu este $0^2+0^2+0^2$, este $0^2+0^2+1^2$

Procedura de calcul ne spune să trecem la următorul număr şi vedem că deşi acesta nu este nici $0^2+0^2+0^2$ şi nici $0^2+0^2+1^2$, este $0^2+1^2+1^2$; trecem acum la 3 şi găsim că $3 = 1^2+1^2+1^2$; apoi la 4, găsind $4 = 0^2+0^2+2^2$ apoi $5 = 0^2+1^2+2^2$; apoi, găsind că $6 = 1^2+1^2+2^2$, trecem la 7, dar acum toate

sumele de trei pătrate mai mici decât 7 şi anume

$0^2+0^2+0^2$ $0^2+0^2+1^2$ $0^2+0^2+2^2$

$0^2+1^2+1^2$ $0^2+1^2+2^2$ $0^2+2^2+2^2$

$1^2+1^2+1^2$ $1^2+1^2+2^2$ $1^2+2^2+2^2$

$2^2+2^2+2^2$

nu dau suma 7, deci calculul se opreşte şi ajungem la concluzia: 7 este un număr dintre cele căutate, nefiind suma a trei pătrate.

2.3. Calcule care nu se termină

După cum s-a văzut, cu procedura de calcul (**A**) am avut noroc. Să presupunem că trebuie să încercăm acum:

(B) Găsiţi un număr care nu este suma a patru pătrate perfecte.

Acum ajungând la 7, găsim că el *este* suma a *patru* pătrate: $7 = 1^2+1^2+1^2+2^2$ deci trebuie să ne mutăm la 8, găsind $8 = 0^2+0^2+2^2+2^2$ apoi la $9 = 0^2+0^2+0^2+3^2$ apoi $10 = 0^2+0^2+1^2+3^2$, etc.

Calculul continuă mai departe şi mai departe:... $23 = 1^2+2^2+3^2+3^2$ $24 = 0^2+2^2+2^2+4^2$ $359 = 1^2+3^2+5^2+18^2$ şi aşa mai departe.

Putem decide că răspunsul la problema pusă este un număr neînchipuit de mare şi deci calculul va dura un timp enorm de lung şi va folosi o enormă memorie pentru a-l găsi. Ar trebui să ne întrebăm dacă există un astfel de număr. Calculul pare să continue şi să nu se mai termine. De fapt, acesta este răspunsul; calculul nu se termină niciodată! Faptul a fost demonstrat de marele matematician francez J.L. Lagrange în 1770 orice număr este suma a patru pătrate perfecte. Nu este o teoremă uşoară (chiar contemporanul lui Lagrange, marele matematician elveţian, Leonhard Euler, un om de o profunzime, originalitate şi productivitate excepţionale a încercat să găsească o demonstraţie, dar nu a reuşit).

Evident, nu voi tulbura cititorul cu detalii ale demonstraţiei lui Lagrange, dar aici voi încerca ceva mult mai simplu:

(C) Să se găsească un număr impar care este suma a două numere pare.

Sper că pentru cititor este clar că acest calcul nu se termină niciodată. Numerele

pare (multiplii lui 2)

0, 2, 4, 6, 8, 10, 12, 14, 16,

dau prin adunare tot numere pare şi deci, cu siguranţă, nu există numere impare, adică unul dintre următoarele numere:

1, 3, 5, 7, 9, 11, 13, 15, 17,...

care să fie suma a două numere pare.

Am dat două exemple (**B**) şi (**C**) de calcule care nu se termină niciodată. În primul caz, proprietatea este mai greu de demonstrat, în vreme ce în al doilea caz, proprietatea calculului de a continua indefinit este aproape evidentă. Să mai dăm un exemplu:

(D) Găsiţi un număr par, mai mare decât 2, care nu este suma a două numere prime.

Reamintim că un număr prim este un număr natural (diferit de 0 sau 1) care nu are alţi factori în afară de 1 şi el însuşi, aşa sunt numerele următoare:

2, 3, 5, 7, 11, 13, 17, 19, 23,...

Este foarte probabil că (**D**) este un calcul care nu se termină niciodată, dar nimeni nu ştie asta cu siguranţă. Acest lucru depinde de adevărul faimoasei „conjecturi a lui Golbach" enunţată de Golbach într-o scrisoare adresată lui Euler în 1742, dar care a rămas nedemonstrată până astăzi.

2.4. Cum să decidem dacă un anumit calcul se termină sau nu ?

Am văzut că un calcul poate să se termine sau nu, şi dacă nu, poate fi simplu să vedem că nu se termină, sau poate fi foarte complicat, sau poate fi atât de complicat încât nimeni nu a fost, până acum, atât de deştept, încât să demonstreze acest lucru.

Prin ce procedeu se conving matematicienii, sau conving pe alţii, că anumite calcule nu se termină de fapt niciodată ? Urmează ei înşişi un procedeu de calcul (un algoritm) pentru a demonstra lucruri de acest fel ? Înainte de a încerca

să răspundem acestei chestiuni, să mai dăm un exemplu, ceva mai greu decât exemplul nostru evident (**C**) dar mult mai simplu decât (**B**). Aici voi încerca să ilustrez modul în care matematicienii îşi obţin uneori concluziile.

Exemplul meu utilizează aşa numitele numere *hexagonale*

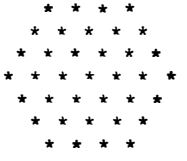
1, 7, 19, 37, 61, 91, 127,...

adică numerele care se pot vizualiza — aranjate ca hexagoane

Aceste numere pot fi obţinute pornind cu 1 şi adunând, succesiv, multiplii lui 6:

6, 12, 18, 24, 30, 36,...

aşa cum se vede, fiecare număr hexagonal poate fi obţinut din cel precedent adăugând câte un inel hexagonal pe margine



şi observând că numărul punctelor în inel trebuie să fie un multiplu de 6, multiplul fiind următorul, de fiecare dată, pe măsură ce hexagonul este mai mare.

Acum să adunăm numerele hexagonale succesive, până la un anumit punct începând cu 1. Ce găsim ?

$1=1$, $1+7=8$, $1+7+19=27$ $1+7+19+37=64$, $1+7+19+37+61=125$

Ce este deosebit în legătură cu numerele 1, 8, 27, 64, 125 ? Toate sunt cuburi. Un cub este un număr care se obţine prin înmulţirea de trei ori a unui număr cu sine însuşi:

$$1=1^3=1\times1\times1,\ 8=2^3=2\times2\times2,\ 27=3^3=3\times3\times3,$$

$$64=4^3=4\times4\times4,\ 125=5^3=5\times5\times5,\ldots$$

Este generală această proprietate pentru numerele hexagonale? Să mai încercăm un caz. Într-adevăr:

$$1+7+19+37+61+91=216=6\times6\times6=6^3.$$

Merge aşa la nesfârşit ?
Dacă da, atunci următoarea procedură de calcul nu se va termina niciodată.

(E) Găsiţi suma numerelor hexagonale succesive, pornind cu 1, care nu este cub.

Am de gând să vă conving că, într-adevăr, acest calcul va continua mereu, fără a se opri.

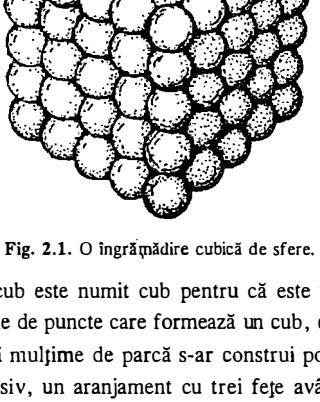


Fig. 2.1. O îngrămădire cubică de sfere.

Mai întâi, un cub este numit cub pentru că este un număr care poate fi reprezentat ca o mulţime de puncte care formează un cub, ca în figura 2.1. Aş dori să vă închipuiţi această mulţime de parcă s-ar construi pornind dintr-un colţ, iar apoi s-ar adăuga succesiv, un aranjament cu trei feţe având fiecare un perete în spate, un perete lateral şi un tavan, ca în figura 2.2.

Să privim acum, acest aranjament cu trei feţe, de undeva de departe, de pe direcţia vârfului comun celor trei feţe. Ce vedem ? Un *hexagon*, ca în figura 2.3. Punctele care constituie aceste hexagoane de dimensiuni tot mai mari, luate împreună, corespund punctelor care formează întregul cub. Acest lucru stabileşte că, adunând succesiv numerele hexagonale, pornind cu 1, vom obţine mereu un

cub. În concluzie am ajuns să arătăm că într-adevăr calculul (E) nu se termină niciodată.

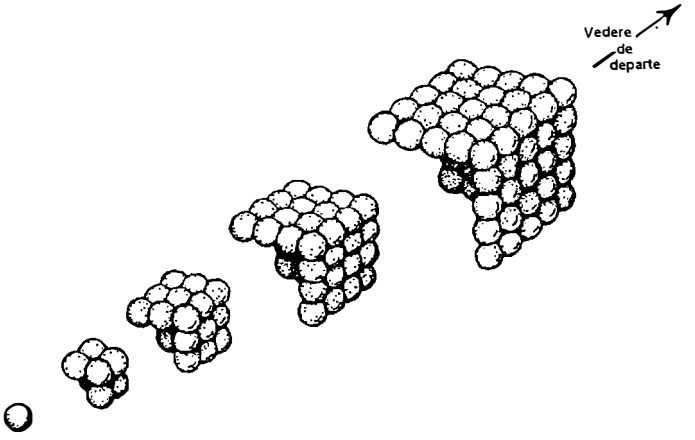


Fig. 2.2. Separarea acestora, fiecare având trei pereți (fundal, lateral și deasupra)

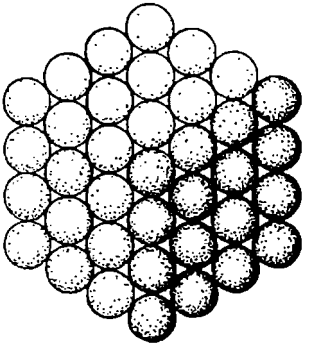


Fig. 2.3. Fiecare bucată este văzută ca un hexagon.

Cititorul ar putea să se teamă că argumentația pe care tocmai am dat-o este într-un anumit fel intuitivă, în loc să fie o demonstrație formală, riguros

matematică. De fapt, argumentaţia este perfect riguroasă şi unul dintre scopurile mele aici este să arăt că există metode riguroase de raţionament matematic care nu sunt 'formalizate', adică nu urmează un sistem de reguli formale. Un exemplu de raţionament geometric mult mai elementar pe care l-am utilizat pentru a obţine o proprietate generală a numerelor naturale este demonstrarea faptului că $a \times b = b \times a$, dată în paragraful 1.19. Şi aceea este o demonstraţie bună, deşi nu este o demonstraţie formală.

Raţionamentul pe care tocmai l-am făcut pentru a însuma numerele hexagonale succesive poate fi eventual înlocuit printr-o demonstraţie mai formală. Partea esenţială a unei astfel de demonstraţii formale se bazează pe principiul inducţiei matematice, care este o metodă de a accede la adevărul unor propoziţii care se aplică *tuturor* numerelor naturale pe baza unui singur calcul. În esenţă, principiul inducţiei matematice ne permite să deducem că o propoziţie $P(n)$, care depinde de un număr natural dat n (ca de exemplu „suma primelor n numere hexagonale este n^3 „), este adevărată pentru *orice* n, cu condiţia să putem arăta mai întâi că ea este adevărată pentru $n=0$, (sau aici $n=1$), iar apoi să putem arăta că adevărul propoziţiei $P(n)$ implică adevărul propoziţiei $P(n+1)$. Nu-l voi plictisi pe cititor cu detaliile demonstraţiei (prin inducţie matematică) faptului că (E) nu se termină niciodată. Cititorul interesat poate, dacă doreşte, să o încerce ca pe un exerciţiu.

Sunt oare regulile bine definite, ca de exemplu principiul inducţiei matematice, totdeauna suficiente pentru a stabili natura infinită (*non-stoping*) a unui calcul care nu se termină niciodată? Răspunsul este, în mod surprinzător „nu" Aceasta este una dintre implicaţiile teoremei lui Gödel, aşa cum vom vedea în curând şi este important să încercăm chiar să o înţelegem. Nu numai principiul inducţiei matematice este insuficient. Dar nici un set de reguli, oricare ar fi ele, nu sunt suficiente pentru a stabili caracterul infinit al unei proceduri de calcul. Prin 'set de reguli' înţelegem un sistem de proceduri formale care pot verifica complet prin calcul dacă au fost sau nu corect aplicate, în orice caz particular. Aceasta pare o concluzie pesimistă, căci rezultă că există calcule care nu se termină niciodată, şi mai mult, faptul că nu se termină niciodată nu poate fi demonstrat riguros matematic. Totuşi, nu acest lucru ni-l spune de fapt teorema lui Gödel. Ceea ce spune teorema lui Gödel poate fi văzut într-o lumină mai bună; şi anume că ceea ce este accesibil matematicienilor — şi de fapt tuturor celor care pot gândi logic

cu înţelegere şi imaginaţie se află undeva deasupra a tot ceea ce poate fi formalizat ca un set de reguli. Regulile pot fi, uneori, un substitut parţial al înţelegerii, dar ele nu pot niciodată să o înlocuiască în întregime.

2.5. Familii de calcule; concluzia Gödel-Turing ($\mathcal{G}$)

Pentru a vedea cum teorema lui Gödel (în forma simplificată pe care o voi da, stimulat de ideile lui Turing) demonstrează cele de mai sus, trebuie să dăm o generalizare a propoziţiilor despre calcul. În loc să ne întrebăm dacă un calcul, ca cel din (**A**), (**B**), (**C**), (**D**) sau (**E**), se termină vreodată sau nu, va trebui să considerăm calcule care să depindă sau să acţioneze asupra unui *număr natural* n. Astfel, dacă notăm un astfel de calcul cu $C(n)$, aceasta înseamnă că va trebui făcut câte un calcul pentru fiecare număr natural 0, 1, 2, 3, 4,..., adică $C(0)$, $C(1)$, $C(2)$, $C(3)$, $C(4)$,... şi că modul în care calculul depinde de n este în sine calculabil.

În termenii maşinilor Turing, $C(n)$ este o acţiune a unei maşini Turing asupra numărului n. Adică, numărul n este introdus de pe banda maşinii ca dată de intrare, iar maşina calculează singură în continuare. Dacă nu vă convine conceptul de 'maşină Turing', gândiţi-vă la un calculator obişnuit, şi luaţi-l pe n ca fiind o valoare iniţială pentru acţiunea unui program ce se execută. Ceea ce ne interesează, pe noi aici, este să stabilim dacă acţiunea calculatorului se încheie pentru fiecare alegere a lui n.

Pentru a lămuri ce se înţelege prin calcul depinzând de un număr natural n, să considerăm două exemple.

(**F**) Să se găsească un număr care să nu fie suma a n pătrate perfecte

şi

(**G**) Să se găsească un număr impar care să fie suma a n numere pare.

Din ceea ce am spus până acum rezultă imediat că (**F**) este un calcul care se termină numai pentru $n = 0$, 1, 2, 3 (caz în care se găsesc respectiv numerele 1, 2, 3, 7). Pe de altă parte, calculul (**G**) nu se termină pentru nici o valoare a lui n. Dacă vrem însă să stabilim că (**F**) nu se termină dacă n este 4 sau mai mare, avem nevoie de o matematică formidabilă (demonstraţia lui Lagrange); în vreme

ce faptul că *(G)* nu se termină niciodată este evident.

Care sunt, în general, metodele care stau la dispoziţia matematicienilor pentru a stabili natura unui calcul? Pot fi puse aceste metode într-o formă calculabilă?

Să presupunem acum că avem un procedeu de calcul A care atunci când se termină ne furnizează demonstraţia faptului că un calcul, precum $C(n)$, de fapt nu se opreşte (ar putea continua indefinit)[1]. Să încercăm să ne imaginăm că A ar încapsula *toate* procedeele disponibile unui matematician pentru a arăta în mod convingător că un anumit calcul poate continua indefinit. Ca urmare, dacă în vreun caz particular A însuşi ajunge cândva la bun sfârşit, acest lucru ne-ar oferi o demonstraţie, făcută de un om, că acel calcul particular la care ne-am referit nu se încheie niciodată. Pentru cele mai multe argumente care vor urma, nu este necesar ca A să aibă acest rol atât de special. Deocamdată ne vom ocupa doar de o parte a raţionamentului matematic. Dar, pentru concluzia noastră finală $\mathcal{G}$, este necesar să ne imaginăm că A încapsulează toate procedurile matematicienilor prin care se poate arăta că un anumit calcul continuă indefinit.

Cu siguranţă, nu voi cere ca A să poată totdeauna decide că $C(n)$ nu se termină atunci când de fapt chiar nu se termină, dar voi insista ca A să nu dea niciodată un răspuns greşit, adică, dacă ajunge vreodată la concluzia că $C(n)$ nu se termină, atunci să fie chiar aşa. Vom spune că A este *corect*[2] dacă nu furnizează răspunsuri greşite.

Să observăm că dacă A nu este corect (spunem ca este *incorect*), atunci acest lucru ar putea fi stabilit, în principiu, cu ajutorul unui calcul direct: Dacă A nu este corect el este verificabil prin calcul. Dacă A ar stabili, în mod eronat, că $C(n)$ nu se opreşte, când de fapt el se termină, atunci executarea efectivă a calculului $C(n)$ ne-ar conduce la concluzia că A trebuie respins (este o altă chestiune dacă un astfel de calcul poate fi practic executat; ea va fi discutată în **Q8**).

Pentru ca A să poată fi aplicat calculului, în general, trebuie să găsim un

---

[1] Pentru obiectivul acestei argumentaţii am adoptat punctul de vedere că, dacă A se termină de tot, atunci acesta este un semnal pentru realizarea unei demonstraţii valabile că $C(n)$ poate continua indefinit. Dacă A 'este în suferinţă' pentru oricare alt motiv decât succesul demonstraţiei, atunci o să considerăm că A nu a reuşit să se termine cu bine. Vezi şi întrebările Q3 şi Q4, precum şi Anexa A

[2] Am tradus termenul englez *sound* prin *corect* (n.t.)

mod de a codifica toate calculele diferite ale lui $C(n)$, astfel încât A să poată utiliza acest cod în acţiunea proprie. Toate variantele posibile de calcul C pot să fie listate, să zicem astfel:

$$C_0,\ C_1,\ C_2,\ C_3,\ C_4,\ C_5,\ldots,$$

iar noi vom spune despre C_q că este calculul al q-ulea. Când un astfel de calcul se aplică unui număr fixat n, vom scrie

$$C_0(n),\ C_1(n),\ C_2(n),\ C_3(n),\ C_4(n),\ C_5(n),\ldots.$$

Putem considera această ordine dată, să zicem, ca pe un anumit tip de ordonare numerică a programelor calculatorului. (Ca să fim mai expliciţi, putem, dacă se cere, să luăm această ordine ca fiind dată de numerotarea maşinilor Turing, aşa cum am descris-o în MN, considerând calculul $C_q(n)$ drept acţiunea executată de maşina Turing T_q asupra lui n). Este important de observat aici un element tehnic şi anume că această listă este *calculabilă*, adică există un singur[3] calcul $C_\bullet$ care ne dă pe C_q când este în prezenţa lui q, sau, mai precis, calculul $C_\bullet$ acţionează asupra perechii de numere q, n (adică q urmat de n) pentru a obţine pe $C_q(n)$.

Procedura A poate fi considerată ca un calcul particular care, atunci când i se prezintă perechea de numere q,n, încearcă să stabilească faptul că $C_q(n)$ este un calcul care ar putea continua indefinit. Astfel, atunci când calculul A se termină, vom avea o demonstraţie a faptului că $C_q(n)$ nu se opreşte. Deşi, aşa cum am spus anterior, vom încerca în curând să ne imaginăm că A ar putea fi formalizarea *tuturor* procedurilor de care dispune un matematician pentru a decide că un calcul nu se termină niciodată, momentan nu este necesar să ne gândim la A în acest mod. Deocamdată, A este o mulţime oarecare de reguli corecte pentru stabilirea faptului că $C_q(n)$ este un calcul care nu se opreşte niciodată. Calculul pe care îl execută A depinde de cele două variabile q şi n şi îl vom nota cu $A(q,n)$; vom avea:

(**i**) Dacă $A(q,n)$ se opreşte, atunci $C_q(n)$ nu se opreşte.

Să considerăm acum propoziţiile particulare (**H**), în care q este luat egal cu n. Pare oarecum straniu să facem aşa ceva, dar este perfect legitim (acesta este primul pas în procedura de diagonalizare, o procedură puternică, descoperită în

[3] De fapt, această unicitate se obţine prin acţiunea maşinii Turing universale asupra perechii de numere q,n ; vezi Anexa A şi MN, capitolul 2.

secolul nouăsprezece de originalul şi influentul matematician danezo-ruso-german, Georg Cantor, centrală atât în raţionamentul lui Gödel cât şi în cel al lui Turing). Luând q egal cu n avem

(I) Dacă $A(n,n)$ se opreşte, atunci $C_n(n)$ nu se opreşte.

Să observăm acum că $A(n,n)$ depinde doar de un singur parametru, şi nu de doi, şi deci trebuie să fie unul dintre calculele C_0, C_1, C_2, C_3, (aplicat lui n), deoarece am presupus că aceasta este lista tuturor calculelor care ar putea fi aplicate asupra numărului natural n. Să presupunem că este vorba de C_k, calcul aplicat lui n, atunci avem:

(J) $A(n,n) = C_k(n)$.

Să examinăm acum valoarea particulară $n=k$ (Aceasta este partea a doua din procedeul de diagonalizare a lui Cantor!) Avem, ţinând seamă de **(J)**,

(K) $A(k,k) = C_k(k)$

şi din (I), cu $n = k$ rezultă:

(L) Dacă $A(k,k)$ se opreşte, atunci $C_k(k)$ nu se opreşte.

Substituind **(K)** în **(L)** găsim:

(M) Dacă $C_k(k)$ se opreşte, atunci $C_k(k)$ nu se opreşte.

De aici ar trebui să deducem că $C_k(k)$ este un calcul care de fapt *nu* se opreşte niciodată. (Căci dacă s-ar opri, atunci nu s-ar opri, în conformitate cu **(M)**!). Dar nici $A(k,k)$ nu poate să se oprească, deoarece din **(K)** rezultă că el coincide cu $C_k(k)$. Astfel, procedura noastră A nu este capabilă să stabilească că un calcul particular $C_k(k)$ nu se opreşte niciodată, nici chiar atunci când el chiar poate continua indefinit.

Mai mult, dacă ştim că A este corect, ştim şi că $C_k(k)$ nu se opreşte. Astfel am ajuns să ştim ceva despre care A nu poate stabili nimic. Rezultă că A nu ar putea cuprinde întreaga noastră înţelegere.

În acest punct un cititor atent s-ar putea să dorească să mai citească odată întregul raţionament, aşa cum a fost prezentat anterior, numai ca să se convingă că eu nu am avut vreo scăpare. Admit că ceva în modul în care se face argumentaţia ar putea aminti de o scamatorie, totuşi, raţionamentul este perfect legitim şi el câştigă în forţă cu cât este examinat mai minuţios. Am găsit deci un calcul $C_k(k)$ despre care ştim că nu se termină, mai mult, procedura de calcul A nu este

suficient de puternică ca să stabilească acest lucru. Aceasta este teorema Gödel (-Turing) în forma de care voi avea nevoie. Ea se aplică la orice procedură de calcul A, oricare ar fi aceasta, pentru a stabili că un calcul nu se termină, *atâta vreme cât ştim că A este corectă*. De aici deducem că nici o mulţime corectă de reguli de calcul, despre care se poate şti că este corectă (aşa cum este A) nu este suficientă pentru a stabili că un calcul nu se termină, deoarece există calcule care nu se termină (ca de exemplu $C_k(k)$) care pot eluda aceste reguli. Mai mult, deoarece din faptul că-l cunoaştem pe A şi din faptul că A este corect, putem construi efectiv un calcul $C_k(k)$ pe care îl *putem vedea* că nu se opreşte , deducem că *A nu poate fi* o formalizare a unei proceduri aflate la dispoziţia matematicienilor pentru a stabili că un calcul nu se termină, indiferent cum ar fi acest A. Deci:

$\mathcal{G}$ **Oamenii care se ocupă cu matematica nu folosesc, pentru a stabili adevărul matematic, vreun algoritm despre care se poate şti că este corect.**

Mi se pare că această concluzie este de neevitat. Totuşi, mulţi au încercat să argumenteze împotriva ei ridicând obiecţii ca cele pe care le vom rezuma în întrebările **Q1–Q20**, din paragrafele 2.6 şi 2.10 care urmează şi desigur mulţi vor obiecta împotriva concluziei mai tari că trebuie să existe ceva fundamental necalculabil în procesele noastre de gândire.

Cititorul ar putea de fapt să se mire de ce oare, un raţionament matematic ca acesta, privind natura abstractă a calculului, ar avea ceva de spus despre modul în care lucrează mintea omului. Şi ce are a face aceasta, cu problema conştiinţei conştiente? Răspunsul este că argumentaţia spune cu adevărat ceva foarte semnificativ despre calitatea mentală a *înţelegerii* în legătură cu problema generală a calculului – şi, aşa cum s-a arătat în paragraful 1.12, calitatea înţelegerii este oarecum dependentă de conştiinţa conştientă. Este adevărat că, în cea mai mare parte, raţionamentele anterioare au fost prezentate utilizând argumente matematice, dar există un punct esenţial şi anume că algoritmul A intervine în argumentare la două niveluri diferite. La un prim nivel el este tratat ca un algoritm oarecare având anumite proprietăţi, dar la un alt nivel, am presupus cu privire la A că el este cu *adevărat* 'algoritmul pe care noi înşine îl folosim' ca să ajungem să credem că un calcul nu se termină. Argumentaţia nu se referă numai la calcule. Ea priveşte şi modul în care este folosită înţelegerea noastră conştientă pentru a deduce adevărul unor propoziţii matematice aici proprietatea lui $C_k(k)$ de a nu

se mai opri. Jocul între aceste două niveluri la care este considerat algoritmul A — ca instanță presupusă a activității conștiente și, în același timp, ca un calcul — ne permite să ajungem la o concluzie care exprimă conflictul fundamental dintre o astfel de activitate conștientă și simplul calcul.

Totuși, există într-adevăr o mulțime de aspecte și contra-argumente care trebuie luate în considerare. În cele ce urmează, voi începe prin a trece în revistă foarte atent *toate* contra-argumentele îndreptate împotriva concluziei $\mathcal{G}$, care mi-au atras atenția acestea vor fi întrebările **Q1** **Q20**, pe care le voi pune în paragrafele 2.6 și 2.10, care includ de asemenea câteva contra-argumente care îmi aparțin. La fiecare dintre aceste întrebări voi răspunde cu atenția cea mai mare de care sunt în stare. Se va vedea că $\mathcal{G}$ rămâne, în urma acestei analize, neatinsă. Apoi, în capitolul 3, mă voi ocupa de implicațiile afirmației $\mathcal{G}$ Vom constata că aceasta furnizează cu adevărat baza pentru un exemplu convingător că înțelegerea matematică conștientă nu poate fi deloc modelată în mod satisfăcător în termeni computaționali, nici în cazul 'algoritmilor predefiniți' (*top-down*), nici în cazul 'algoritmilor restructurabili' (*bottom-up*) și nici în cazul vreunei combinații a lor. Mulți ar putea considera aceasta ca o concluzie alarmantă, deoarece pare că ea ne lasă fără spațiu de mișcare.

În partea a doua a acestei cărți, vom porni pe o direcție mai constructivă. Voi dezvolta ceea ce eu cred că ar putea fi o construcție științifică plauzibilă pentru speculațiile mele proprii asupra proceselor fizice despre care s-ar putea concepe că stau la baza activității creierului. De exemplu, ce se întâmplă atunci când urmărim o argumentație de acest fel și modul în care aceste procese ar putea scăpa oricărei descrieri computaționale.

2.6. Obiecțiile tehnice posibile relative la $\mathcal{G}$

Concluzia $\mathcal{G}$ poate fi resimțită de cititor ca una surprinzătoare, mai ales dacă ținem seamă de faptul că ingredientele argumentelor prin care a fost dedusă sunt relativ simple. Înainte de a trece, în capitolul 3, la implicațiile lui $\mathcal{G}$, implicații cu privire la posibilitatea de a construi un robot controlat de un calculator, inteligent și în stare să facă matematică, trebuie să examinăm, cu multă atenție, câteva aspecte tehnice care privesc deducerea lui $\mathcal{G}$. Dacă sunteți

un cititor mai puţin interesat de anumite artificii tehnice şi sunteţi gata să acceptaţi concluzia $\mathcal{G}$ – că matematicienii nu folosesc un algoritm despre care să poată şti că este corect pentru a stabili adevărurile matematice – atunci puteţi sări aceste argumente (cel puţin pe moment) şi să treceţi direct la capitolul 3; ba mai mult, dacă sunteţi pregătit să acceptaţi concluzia încă mai tare, că nu poate exista nici o explicaţie algoritmică pentru înţelegerea matematică, sau de altă natură, atunci s-ar putea să preferaţi să treceţi direct la partea a doua, făcând, eventual o pauză, pentru a examina dialogul fantastic din paragraful 3.23 (care rezumă argumentele esenţiale din capitolul 3) şi a parcurge concluziile din paragraful 3.28.

Legat de metoda matematică folosită, există câteva puncte care-i pot nelinişti pe unii, cu privire la raţionamentul de tip Gödel folosit în paragraful 2.5. Să încercăm în continuare să le clasificăm.

Q1. L-am considerat pe A ca fiind o singură procedură, când fără îndoială utilizăm mai multe tipuri de raţionamente într-o demonstraţie matematică. Nu ar fi fost necesară o întreagă listă de A-uri?

De fapt, în modul în care am tratat eu problema, nu am pierdut nimic din generalitate. Orice listă finită A_1, A_2, A_3, ..., A_r de proceduri algoritmice poate fi exprimată ca un singur algoritm A, în aşa fel încât A să nu se termine numai dacă nici una dintre procedurile A_1, A_2,..., A_r nu se termină. (Procedura A ar putea fi descrisă, în cazul acesta, astfel „ Execută primii 10 paşi din procedura A_1 memorează rezultatul; execută primii 10 paşi din procedura A_2 memorează rezultatul; execută primii 10 paşi din procedura A_3; memorează rezultatul; şi aşa mai departe, până la A_r; apoi reia procedura A_1 şi execută următorii săi 10 paşi; memorează rezultatul; şi aşa mai departe; apoi treci la a treia secvenţă de câte 10 paşi, etc. Opreşte-te imediat ce se termină vreuna dintre procedurile A_r „.) Dacă, pe de altă parte, lista procedurilor lui A ar fi infinită, atunci, pentru a putea fi considerată o procedură algoritmică, ar trebui să existe un mod de a genera în întregime mulţimea A_1, A_2, A_3, ...pe baza unui algoritm. Vom obţine astfel algoritmul A, în locul listei iniţiale cu următorul procedeu:

primii 10 paşi ai lui A_1;

seria a doua de 10 paşi din A_1; primii 10 paşi din A_2;

seria treia de 10 paşi din A_1; seria doua de 10 paşi din A_2; primii 10 paşi din A_3;

...etc...

Algoritmul se termină de îndată ce una dintre procedurile listei se încheie cu succes.

Am putea, pe de altă parte să ne imaginăm că lista $A_1, A_2, A_3, \ldots$, pe care o considerăm infinită, să nu poată fi enumerată dinainte, nici măcar în principiu. Putem să ne imaginăm că proceduri algoritmice succesive pot fi adăugate din timp în timp listei, fără ca inițial acest lucru să fi fost specificat complet. În orice caz, în absența unei proceduri algoritmice de generare a listei, specificată în prealabil, nu avem de fapt o procedură auto-conținută.

Q2. Oare nu s-ar putea să permitem ca algoritmul A să nu fie fixat ? Ființele umane pot învăța; deci algoritmul pe care ființele umane îl folosesc ar putea fi unul care se schimbă continuu.

Un algoritm care se schimbă ar avea nevoie de unele specificații privind regulile după care au loc schimbările. Dacă aceste reguli sunt pe de-a-ntregul algoritmice, atunci ele ar fi trebuit să fie deja incluse în ceea ce am considerat a fi A; astfel, acest tip de 'algorim de schimbare' este de fapt doar o altă instanțiere a unui anumit algoritm și argumentația poate decurge ca mai sus. Pe de altă parte, ne putem imagina moduri în care algoritmul s-ar putea schimba care să nu fie algoritmice și, s-ar putea sugera, pentru o astfel de situație, încorporarea unor ingredienți aleatori sau anumite tipuri de interacțiuni cu mediul. Statutul de 'nealgoritmicitate' ale unor asemenea moduri de schimbare ale unui algoritm, va fi tratat mai târziu (în paragrafele 3.9 și 3.10); vezi de asemenea discuția din paragraful 1.9, unde se prezintă argumentele după care nici aceste metode nu furnizează un mijloc de a evada din algoritmism[4] (așa cum s-ar cere din punctul de vedere $\mathcal{C}$). Pentru scopurile noastre de moment, pur matematice, ne vom ocupa numai de posibilitatea în care schimbarea este cu adevărat algoritmică. Pentru că, odată ce am acceptat că astfel de schimbări nu pot fi algoritmice, ajungem la aceeași concluzie ca și $\mathcal{G}$.

Sigur, ar trebui să fiu ceva mai explicit în legătură cu ceea ce s-ar putea înțelege printr-un algoritm A care se schimbă algoritmic. Putem presupune că A depinde nu numai de q și n, ci de încă un parametru t, la care ne-am putea gândi că reprezintă 'timpul', sau poate mai bine că t numără de câte ori

[4] Cuvântul *algoritmism* , definind, în esență, 'punctul de vedere $\mathcal{A}$' a fost introdus de Hao Wang în 1993.

algorimul a fost activat mai înainte. În ori ce caz, putem presupune că parametrul t este un număr natural. Avem astfel algoritmii $A_t(q,n)$ care pot fi listaţi:

$$A_0(q,n),\ A_1(q,n),\ A_2(q,n),\ A_3(q,n),$$

unde se presupune că fiecare este o procedură validă pentru a stabili că $C_q(n)$ este un calcul care nu se termină, dar unde ne imaginăm că acestor proceduri li se permite să crească în putere pe măsură ce t creşte. Metoda prin care ele cresc în putere se presupune că este algoritmică. Probabil că această 'metodă algoritmică' este ceva care ar putea să depindă de 'experienţa' algoritmilor $A_t(q,n)$ precedenţi, dar această 'experienţă' este aici considerată ca fiind generată algoritmic (căci altfel ajungem din nou de acord cu $\mathcal{G}$) şi astfel va trebui să îi includem pe ei, sau metoda lor de generare, în ceea ce va constitui următorul algoritm (adică în $A_t(n,q)$ însuşi). În acest fel, ajungem la un *singur* algoritm $(A_t(q,n))$ care depinde algoritmic de trei parametri t, q, n. De aici putem construi un algoritm A^* care să fie tot atât de puternic ca întreaga listă $A_t(q,n)$, dar care să depindă numai de doi parametri naturali q şi n. Pentru a-l construi pe $A^*(q,n)$ ca mai sus, tot ceea ce ne trebuie este să-l lăsăm să funcţioneze şi să parcurgă primii 10 paşi ai lui $A_0(q,n)$, să memoreze rezultatul; apoi primii 10 paşi ai lui $A_1(q,n)$ urmaţi de tranşa a doua de 10 paşi ai lui $A_0(q,n)$, memorând rezultatele; apoi primii 10 paşi ai lui $A_2(q,n)$, a doua tranşă de 10 paşi ai lui $A_1(q,n)$ şi a treia tranşă de 10 paşi ai lui $A_0(q,n)$; etc., unde la fiecare etapă rezultatele sunt memorate; în final ajungem să ne oprim de îndată ce vreuna dintre componentele calculului se termină. Utilizând A^* în locul lui A, raţionamentul care susţine pe $\mathcal{G}$ poate fi condus ca mai înainte.

Q3. **N-am fost oare, în mod inutil, prea restrictiv când am insistat că A trebuie să continue indefinit să calculeze în acele cazuri în care devine clar că $C_q(n)$ de fapt se opreşte? Dacă am permite ca A să se termine în acele cazuri, argumentaţia noastra ar eşua. Discernământul pe care fiinţele umane îl posedă, cu siguranţă că le permite, uneori, să tragă concluzia că un calcul se termină, dar eu las impresia că aş ignora aceste lucruri. Înseamnă asta oare că am fost prea restrictiv?**

Nu, deloc. Argumentaţia noastră se presupune că se aplică numai la capacitatea noastră de a înţelege că un anumit calcul nu se termină, şi nu la acele intuiţii care ne-ar permite să tragem concluzia opusă. Algoritmului considerat, A, nu i se

permite să ajungă la 'o încheiere cu succes' pe baza concluziei că un anumit calcul se termină. Acest lucru nu face parte din sarcinile sale.

Dacă nu vă prea place acest lucru, gândiţi-vă la A în următorul fel: încercaţi să includeţi în A ambele capacităţi de înţelegere, dar cu precizarea că atunci când calculele $C_q(n)$ se termină, algoritmul A să intre într-un ciclu infinit (de exemplu, să repete din nou şi din nou o aceeaşi operaţie fără să se mai oprească). Evident, nu acesta este modul în care operează un matematician, dar până la urmă asta nu are nici o importanţă. Argumentaţia noastră are aici forma unei reduceri la absurd, şi anume, pornim de la presupunerea că noi utilizăm un algoritm corect pentru a ajunge la adevărul matematic şi apoi demonstrăm că ajungem la o contradicţie. În acest raţionament nu este necesar ca A să fie cu adevărat chiar algoritmul presupus, dar trebie să fie unul realizat pornind de la A, aşa cum l-am descris mai înainte.

Acelaşi comentariu se poate aplica oricărei alte obiecţii aduse argumentaţiei din paragraful 2.5 de forma „Cu siguranţă A s-ar putea opri pentru tot felul de motive false fără ca astfel să furnizeze demonstraţia că $C_q(n)$ nu s-ar termina" Dacă ni se dă un A care se comportă în acest fel, aplicăm pur şi simplu argumentaţia din paragraful 2.5 unui algoritm A modificat, şi anume unul care intră într-un ciclu infinit ori de câte ori algoritmul original A se opreşte, pentru oricare dintre motivele false evocate anterior.

Q4. **Numerotând C_0, C_1, C_2, s-ar părea că am presupus că fiecare C_q se referă la un calcul bine definit; în timp ce în orice tip de ordonare numerică sau lexicografică a programelor de calcul, lucrurile, cu siguranţă, nu stau aşa ?**

Ar fi cu adevărat lipsit de eleganţă să presupun că numerotarea pe care am făcut-o ar furniza un calcul efectiv C_q, pentru fiecare număr natural q. De exemplu, numerotarea maşinilor Turing T_q dată în MN nu realizează, cu siguranţă, o astfel de situaţie; vezi p. 60-67. Pentru un q fixat, maşina Turing T_q , aşa cum a fost descrisă acolo, ar trebui considerată 'un rebut' pentru unul dintre următoarele motive: s-ar putea să ruleze indefinit, fără să se oprească vreodată; s-ar putea să nu fi fost 'corect specificată', deoarece numărul n în binar, are prea multe cifre *1* în dezvoltarea sa (cinci sau mai multe valori 1 succesive) şi astfel să nu existe nici o translatare în schema dată; s-ar putea să dea peste o instrucţiune care să introducă o stare internă inexistentă sau s-ar putea să producă pe bandă numai

spaţii goale când se opreşte, fapt care nu are nici o interpretare (vezi de asemenea Anexa A).

Din punctul de vedere al raţionamentului Gödel pe care tocmai l-am dat, tot ceea ce trebuie să facem este să îngrămădim toate aceste situaţii împreună, sub eticheta 'nu se opreşte'. În particular, când am spus 'se termină' relativ la procedura de calcul A, (vezi şi nota de subsol nr. 2, din acest capitol), acest lucru implică faptul că cu adevărat 'se opreşte' în sensul celor spuse anterior (şi astfel nu conţine şiruri netranslatabile şi nici nu produce numai spaţii goale pe bandă) — adică 'se opreşte' implică faptul că este vorba despre un calcul efectiv, cu adevărat bine specificat. De asemenea, '$C_q(n)$ se opreşte' mai înseamnă că, de fapt, calculul se opreşte în mod corespunzător.

Cu această interpretare, argumentaţia, aşa cum am dat-o nu este afectată de obiecţia Q4.

Q5. Nu-i aşa că am arătat doar că este posibil să depăşim o procedură algoritmică particulară A, făcând-o să eşueze cu ajutorul calculului $C_q(n)$? De ce ar arăta asta că eu pot face mai bine decât oricare A ?

Argumentaţia arată că noi, cu siguranţă, putem face mai bine deducţia decât oricare algoritm. Acesta este punctul central al raţionamentului prin reducere la absurd pe care l-am folosit aici.

Mă gândesc că, în acest moment, o analogie ar putea fi de folos. Unii dintre cititori au întâlnit în şcoală afirmaţia lui Euclid care spune că nu există un cel mai mare număr prim. Demonstraţia se face, de asemenea, prin reducere la absurd. O prezentăm în cele ce urmează. Să presupunem prin absurd că există un cel mai mare număr prim, să-l notăm cu p. Apoi, să considerăm produsul N al tuturor numerelor prime până la p la care să mai adunăm pe 1

$$N = 2\times3\times5\times\ldots\times p+1.$$

N este cu siguranţă mai mare decât p, dar nu se divide prin nici unul dintre numerele prime 2, 3, 5, ..., p (deoarece prin împărţire rămâne restul 1); rezultă deci că, sau N este numărul prim căutat, sau el este un număr compus, caz în care trebuie să se dividă printr-un număr mai mare decât p. În ambele cazuri rezultă că există un număr prim mai mare decât p, ceea ce contrazice presupunerea că p ar fi numărul prim cel mai mare. Deci nu există un cel mai mare număr prim.

Demonstrația prin reducere la absurd nu arată numai că un anumit număr prim, p, poate fi întrecut, găsind un număr prim mai mare; ea arată că nu există nici un număr prim mai mare decât toate celelalte. La fel, raționamentul Gödel-Turing dat anterior, nu arată numai faptul că un anumit algoritm A poate fi întrecut, ci arată că nu există nici un algoritm despre care se poate ști că este corect, care să fie echipat cu discernământul pe care îl folosim pentru a stabili că un anumit calcul nu se oprește.

Q6. Un calculator poate fi programat să urmeze cu precizie argumentația pe care am dat-o aici. N-ar putea el însuși să ajungă la concluzia la care am ajuns și eu ?

Este adevărat că, dat fiind algoritmul A, găsirea calculului particular $C_k(k)$ este un proces de calcul. Și acest lucru poate fi expus explicit[5] Asta înseamnă că intuiția matematică pe care am presupus-o nealgoritmică – intuiția care ne permite să apreciem faptul că $C_k(k)$ nu se termină niciodată – este de fapt algoritmică ?

Cred că trebuie să acordăm acestei discuții o anumită atenție pentru că ea privește una dintre cele mai răspândite concepții greșite în legătură cu demonstrația lui Gödel. Trebuie să fie clar că ea nu invalidează nimic din ceea ce nu era mai înainte. Deși procedura de obținere a lui $C_k(k)$ din A poate fi pusă sub forma unui calcul, acest calcul nu face parte dintre procedurile pe care le conține A. Și nici nu poate face parte, deoarece A nu este capabil să stabilească adevărul lui $C_k(k)$, pe când acest nou calcul (împreună cu A) se presupune că o poate face. Astfel, cu toate că noul calcul este cu adevărat un calcul care conduce la $C_k(k)$, el nu este dintre acelea care au fost admise în clubul celor care 'stabilesc adevărul oficial'.

Permiteți-mi să prezint lucrurile și într-un alt mod. Să ne imaginăm un robot controlat de un calculator care este în stare să stabilească adevăruri matematice cu ajutorul procedurilor conținute în A. Pentru a face lucrurile mai plastice voi utiliza o terminologie antropomorfică și voi spune că robotul 'cunoaște' adevărul matematic – aici infinitatea calculului care poate fi dedusă din utilizarea lui A. Totuși, dacă A este tot ceea ce robotul 'știe', atunci el nu va 'ști' că $C_k(k)$ nu se termină, chiar dacă procedura de obținere a lui $C_k(k)$ din A

---

[5] Pentru a accentua faptul că prețuiesc acest punct, trimit cititorul la Anexa A, unde este expus un procedeu de calcul (urmând regulile explicate în detaliu în MN, Capitolul 2) pentru a obține din algoritmul A, acțiunea $C_k(k)$ asociată unei mașini Turing. Aici A se presupune că este dat de mașina Turing T_a, a cărei evaluare pentru $C_q(n)$ este codificată ca o acțiune a lui T_a pe q urmat de n

este perfect algoritmică. Evident, putem să-i *spunem* robotului nostru că $C_k(k)$ nu se termină cu adevărat (bazându-ne pentru asta pe propria intuiţie), dar dacă robotul ar accepta acest lucru, el ar trebui să-şi modifice propriile reguli, adăugând acest nou adevăr celor pe care deja le 'ştie'. Ne putem imagina, mai departe, că spunem robotului în mod potrivit că procedura generală de calcul pentru obţinerea lui $C_k(k)$ din A este ceva ce s-ar cuveni să 'ştie', ca mod de a obţine adevăruri noi din cele vechi. Orice fel de lucruri care sunt bine definite şi calculabile pot fi adăugate setului de cunoştinţe ale robotului. Dar acum avem un 'A' nou şi raţionamentul lui Gödel poate fi aplicat acestuia în locul lui A, cel vechi. Adică, vom utiliza acest nou 'A', peste tot în locul lui A, cel vechi, deoarece ar fi incorect să-l schimbăm pe 'A' în mijlocul argumentaţiei. Astfel, vedem că ceea ce nu merge cu **Q6** seamănă foarte bine cu ceea ce nu mergea cu **Q5**, aşa cum am discutat mai înainte. În raţionamentul prin reducere la absurd presupunem că A care trebuie să fie o procedură corectă, dar necunoscută, care stabileşte că un calcul nu se termină reprezintă, de fapt, totalitatea procedurilor aflate la dispoziţia matematicienilor şi, de aici, rezultă o contradicţie. Este greşit să introducem o altă procedură de calcul pentru validarea adevărului, care să nu fie conţinută în A, după ce am presupus că A reprezintă totalitatea procedurilor de acest tip.

Problema neplăcută pentru bietul nostru robot este că, în absenţa oricărei *înţelegeri* a procedurii lui Gödel, robotul nu are nici un mijloc independent, demn de încredere să judece adevărul, altul decât cel pe care i-l spunem noi (Aceasta este o chestiune diferită de aspectele computaţionale din raţionamentul lui Gödel). Pentru a fi capabil să facă mai mult decât atât, el, la fel cu noi, ar trebui să înţeleagă semnificaţia operaţiilor care îi sunt indicate. Fără înţelegere, el poate, la fel de bine (sau de rău), să 'ştie' că $C_k(k)$ se termină, în loc să 'ştie' că nu se termină. Este doar o chestiune de algoritm să deduci (în mod greşit) că '$C_k(k)$ se termină', sau să deduci (corect) că '$C_k(k)$ nu se termină' Rezultă că natura algoritmică a acestor operaţii este ne-esenţială; este esenţial că robotul nostru cere judecăţi valide asupra adevărului, pentru a şti care algoritmi îi dau adevăruri şi nu falsuri.

Deci, în acest punct al argumentaţiei, este încă posibil ca 'înţelegerea' să fie un alt tip de activitate algoritmică, neconţinută în nici o procedură corectă cunoscută, dată în mod precis, ca A De exemplu, înţelegerea ar putea fi dată

printr-un algoritm incorect sau necognoscibil.

Într-o discuţie viitoare (în capitolul 3), voi încerca să-l conving pe cititor că, de fapt, înţelegerea nu este o activitate algoritmică. Dar deocamdată, ne preocupăm de implicaţiile raţionamentului Gödel — Turing şi, pentru acest lucru, faptul că $C_k(k)$ poate fi obţinut din A într-un mod calculabil, nu este cu adevărat important.

Q7. **Toate cunoştinţele tuturor matematicienilor care au trăit vreodată, împreună cu toate cunoştinţele matematicienilor din următorii (să zicem) 1000 de ani, sunt în număr finit şi pot fi memorate într-o bancă de date a unui calculator potrivit ales. Oare acest calculator special ar *putea* să simuleze aceste cunoştinţe comportându-se (din punct de vedere extern) ca un om care face matematică, în ciuda faptului că raţionamentul Gödel ne spune chiar contrariul?**

Câtă vreme presupunem că acest lucru este adevărat, ignorăm faptul esenţial şi anume modul în care, noi (sau calculatorul) ştim care afirmaţii — propoziţii matematice — sunt adevărate şi care false (în orice caz, memorarea afirmaţiilor matematice este un lucru ce poate fi realizat cu un sistem mult mai puţin sofisticat decât un calculator, de exemplu în mod fotografic).

Modul în care calculatorul urmează să fie folosit, conform cu Q7, ignoră elementul critic şi anume judecata asupra adevărului. Se pot imagina, la fel de bine, calculatoare care ar conţine doar 'teoreme' matematice complet false, sau liste care să conţină, în mod aleator şi de-a-valma, propoziţii false şi adevărate. Cum am putea spune în care calculator să te încrezi? Ceea ce încerc să explic aici nu este faptul că o simulare a activităţii umane conştiente (aici matematice) este imposibilă, ci, eventual, că din pură întâmplare, calculatorul ar putea să aibă dreptate — chiar fără nici o înţelegere. Dar şansele împotriva acestei eventualităţi sunt absurd de mari şi chestiunile discutate aici — cum decidem care afirmaţii matematice sunt adevărate şi care false — nici măcar nu au fost atinse de Q7.

Există pe de altă parte, un aspect mult mai serios care este atins cu adevărat în Q7. Este întrebarea dacă discuţiile despre structuri infinite (de exemplu, toate numerele naturale, sau toate calculele) sunt relevante pentru consideraţiile prezente, când output-urile calculatoarelor şi oamenilor sunt finite. Vom considera acest aspect important în mod separat în cele ce urmează.

Q8. Calculele care nu se termină niciodată sunt construcţii matematice ideale care operează cu 'infinitul'. Dar sunt astfel de chestiuni relevante în discuţiile despre obiecte fizice finite, aşa cum sunt calculatoarele şi creierele?

Este evident că, în discuţiile noastre idealizate asupra maşinilor Turing sau a calculelor care se pot continua indefinit, am considerat procese (potenţial) infinite, în vreme ce, atunci când e vorba de oameni şi calculatoare, avem de a face cu sisteme finite. Este important să încercăm să stabilim limitele unor astfel de argumente idealizate atunci când le aplicăm unor obiecte fizice finite. Însă, considerarea finitudinii nu afectează în mod substanţial raţionamentul Gödel-Turing. Nu este nimic incorect în discutarea calculelor idealizate, în raţionamentul asupra lor şi în deducerea, matematică, a limitărilor lor teoretice. Putem, de exemplu, în termeni finiţi să discutăm dacă există sau nu un număr impar care să fie suma a două numere pare, sau dacă există un număr natural care nu este suma a patru pătrate (ca în (**C**) şi (**B**)), în ciuda faptului că în abordarea acestei chestiuni considerăm în mod implicit mulţimea infinită a *tuturor* numerelor naturale. Putem să raţionăm cu toată îndreptăţirea asupra calculelor care nu se termină niciodată sau, în general, asupra maşinilor Turing, considerate construcţii matematice, chiar dacă nu este posibilă construcţia efectivă, în practică, a unei maşini Turing cu o bandă fără sfârşit (în particular, să observăm că acţiunea unei maşini Turing care caută un număr impar care să fie suma a două numere pare nu poate fi fizic implementată; căci mai degrabă s-ar uza componentele maşinii decât să funcţioneze la nesfârşit). Specificarea oricărui calcul dat (sau acţiune a unei maşini Turing) este totdeuna finită şi întrebarea dacă, eventual, acţiunea se va termina, este perfect definită. După ce raţionamentul nostru asupra unor astfel de calcule idealizate se încheie, abia *atunci* încercăm să vedem în ce mod discuţia noastră se poate aplica sistemelor finite, ca de exemplu calculatoare sau oameni.

Limitările legate de caracterul finit pot să apară în două cazuri:

(i) atunci când specificarea unui calcul este enorm de mare (adică, numărul n din C_n sau perechea de numere q,n care intervine în $C_q(n)$, sunt prea mari pentru a fi acceptate de un calculator funcţional sau de o persoană),

(ii) atunci când calculul, nefiind prea mare poate fi specificat, dar durează prea mult pentru a fi executat, încât ar putea să pară că nu se va termina niciodată, cu toate că, teoretic, se poate arăta că se termină.

De fapt, aşa cum vom vedea imediat, dintre cele două variante, doar (i) intervine în discuţia noastră, deşi nici chiar (i) nu o afectează foarte mult. Faptul că (ii) nu este important s-ar putea să pară surprinzător. Există o mulţime de calcule simple, care, în cele din urmă se opresc, dar, pentru care nici un calculator imaginabil nu poate calcula destul pentru a ajunge la oprire. Ca exemplu, să luăm următoarea instrucţiune:

„tipăreşte cifra 1 de $2^{2^{65536}}$ ori şi apoi opreşte-te"

(vom mai da o mulţime de exemple interesante în paragraful 3.26.). Dacă un calcul se va opri sau nu, aceasta e o problemă care nu trebuie tratată direct, calculând; o astfel de metodă este total ineficientă. Pentru a vedea cum limitările privind caracterul finit (i) şi (ii) pot afecta discuţia noastră, să reluăm partea esenţială din argumentaţia de tip Gödel. În conformitate cu limitarea (i), în locul listei infinite de calcule vom avea o listă *finită*:

$$C_0,\ C_1,\ C_2,\ C_3,\ldots,\ C_Q,$$

unde, Q este cel mai mare număr care specifică calculul pe care calculatorul (sau o persoană) îl poate gestiona. În cazul unei fiinţe umane, evident precizarea acestui număr este cam vagă. Pentru moment, nu-i important ca Q să fie dat (problema vagului în aprecierea capacităţilor umane va fi discutată mai târziu în răspunsul la **Q13**, în paragraful 2.10). Mai mult, putem presupune că, atunci când aplicăm aceste calcule asupra unui număr natural n, valoarea lui n nu va depăşi un număr N fixat şi suficient de mare, deoarece calculatorul (ca şi fiinţa umană) nu este capabil să mânuiască numere mai mari decât N (mai precis, am putea să luăm în considerare posibilitatea ca N să nu fie un număr fixat, ci să depindă de calculul particular C_q pe care-l considerăm — adică N ar putea să depindă de q ; acest lucru nu aduce, însă, nimic nou în consideraţiile noastre).

Să considerăm, deci, ca mai înainte, un algoritm corect, $A(q,n)$, care, atunci când se încheie, ne furnizează demonstraţia faptului că $C_q(n)$ este un calcul care nu se opreşte. Când vorbim despre un algoritm 'corect', deşi în conformitate cu (i), trebuie să considerăm doar valori ale lui q care nu sunt mai mari decât Q şi valori ale lui n care nu sunt mai mari decât N, de fapt înţelegem că A este corectă pentru orice valoare q şi n, oricât de mari ar fi ele (astfel, regulile incluse în A sunt reguli matematice precise şi nu aproximative, care funcţionează în virtutea unor limitări practice asupra calculelor care pot fi efectuate, de fapt).

Mai mult, când spunem ' $C_q(n)$ nu se opreşte' înţelegem că în *realitate* nu se termină şi nu că ar putea dura prea mult calculul, pentru a putea fi executat de un calculator sau fiinţă umană, aşa cum este prevăzut în (ii).

Să ne amintim că (**H**) spune:

Dacă $A(q,n)$ se opreşte, atunci $C_q(n)$ nu se opreşte.

Dacă ne gândim la (ii), am putea considera că algoritmul A nu ne este de mare folos, pentru a decide dacă un alt calcul nu reuşeşte să se termine, dacă el însuşi are mai mulţi paşi de executat decât poate calculatorul nostru (sau omul) să execute. Dar acest lucru nu are importanţă pentru argumentaţia noastră. Vom căuta un calcul $A(k,k)$ care nu se termină deloc. Nu are importanţă pentru noi că, în anumite cazuri, A se opreşte de fapt, dar noi nu putem aştepta suficient de mult ca să constatăm acest lucru.

Acum, ca în (**J**), vom determina un număr natural k pentru care calculul $A(n,n)$ este acelaşi ca şi $C_k(n)$ pentru orice n:

$$A(n,n) = C_k(n)$$

Trebuie să considerăm posibilitatea ca acest k să fie mai mare decât Q, aşa cum se consideră în (i). Aşa s-ar putea întâmpla atunci când A ar fi extraordinar de complicat şi s-ar apropia de limita de sus (în termenii numărului de cifre binare din specificarea maşinii Turing corespunzătoare) pe care calculatorul sau omul o poate gestiona. Aceasta se întâmplă deoarece calculul prin care se obţine k din specificarea lui A (ca maşină Turing) este destul de simplu şi poate fi dat explicit (aşa cum am arătat deja în răspunsul la **Q6**). Calculul de care avem nevoie pentru a-l 'înfrânge' pe A este $C_k(k)$ şi punând $n = k$ în (**H**) obţinem (**L**)

Dacă $A(k,k)$ se opreşte, atunci $C_k(k)$ nu se opreşte.

Cum $A(k,k)$ coincide cu $C_k(k)$, demonstraţia noastră arată că $C_k(k)$ este un calcul care nu se poate termina de loc, deşi A nu poate stabili acest fapt, chiar dacă ar putea rula mult mai mult timp decât limita care se impune în concordanţă cu (ii). Specificarea lui $C_k(k)$ se dă, în termenii descrişi anterior, pentru k şi cu precizarea că k nu este mai mare, nici decât Q, nici decât N, adică, este un calcul ce poate fi efectiv implementat într-un calculator (sau într-o fiinţă umană) — în sensul că acest calcul poate *începe*. Dar, în nici un caz, calculul nu poate continua până la terminare, pentru că, de fapt, este un calcul care nu se opreşte niciodată!

Dar acum, poate k să fie mai mare decât Q sau N? Aşa s-ar întâmpla

doar dacă specificarea lui A ar cere atât de multe cifre încât, mărind acest număr cu puţin, am fi conduşi la depăşirea capacităţii calculatorului (sau a posibilităţii de calcul a fiinţei umane). Din faptul că A este un algoritm corect rezultă că noi *ştim* că acest $C_k(k)$ nu se poate termina chiar dacă avem dificultăţi în implementarea efectivă a calculului $C_k(k)$ Presupunerea (i) ne impune să luăm în considerare posibilitatea ca A să fie un calcul atât de complicat, încât specificarea lui să-l ducă spre limita posibilităţilor de calcul ale fiinţei umane, şi doar o creştere relativ mică în numărul cifrelor considerate să conducă la un calcul ce depăşeşte capacitatea umană. Mă gândesc însă că, oricum am privi o astfel de posibilitate, oricare ar fi setul de reguli de calcul incluse în A, ele ar fi atât de complicate încât corectitudinea algoritmului nu ar putea fi stabilită în mod plauzibil, chiar dacă regulile ar fi cunoscute. Astfel concluzia noastră este aceeaşi ca mai înainte: nu accedem la adevărul matematic cu ajutorul unei mulţimi de reguli algoritmice a căror corectitudine poate fi cunoscută.

Merită sa fim ceva mai exacţi în privinţa dificultăţii trecerii de la A la $A_k(k)$. Acest lucru are o semnificaţie deosebită pentru noi în cele ce urmează (paragraful 3.19 şi 3.20). În Anexa A, la sfârşitul acestui capitol, se dă o specificare explicită pentru $C_k(k)$ în termenii unei maşini Turing, aşa cum a fost ea descrisă în **MN** capitolul 2. În conformitate cu acea descriere, cu T_m am notat „a m-a maşină Turing" Aici, pentru claritate, voi folosi aceeaşi notaţie, în particular pentru a defini *gradul de dificultate* al unui procedeu de calcul sau al unui calcul individual. Vom defini deci gradul de dificultate μ al maşinii Turing T_m ca numărul de cifre din reprezentarea binară a numărului m (în conformitate cu **MN**, capitolul 2.); atunci gradul de dificultate al unui calcul particular $T_m(n)$ este definit ca cel mai mare dintre cele două numere μ şi ν, unde cu ν am notat numărul cifrelor din reprezentarea binară a lui n Acum să considerăm prescripţia explicită de obţinere a calculului $C_k(k)$ în termenii unei maşini Turing (vezi Anexa A). Dacă presupunem că A are gradul de dificultate α atunci găsim că gradul de dificultate explicit $C_k(k)$ este mai mic decât

$$\alpha+210\times\log_2(\alpha+336),$$

un număr mai mare decât α, cu o valoare relativ mică, atunci când α este foarte mare.

Există în linia generală a argumentaţiei anterioare o clauză care ar putea nelinişti pe unii din cititori. Are sens să considerăm un calcul care ar putea fi prea

complicat pentru a putea fi scris, sau dacă ar fi scris, executarea lui ar dura mult mai mult decât vârsta actuală a universului, chiar dacă fiecare pas al calculului ar fi executat în cea mai mică fracţiune de secundă în care un proces fizic poate fi imaginat a avea loc ? Calculul considerat anterior care tipăreşte succesiunea de

$$2^{2^{65536}}$$

cifre 1 şi care se opreşte după ce această sarcină s-a încheiat, este un astfel de exemplu şi ar fi extrem de neconvenţional, din punct de vedere matematic, să ne permitem să spunem că acesta este un calcul care nu se opreşte. Totuşi, există unele puncte de vedere matematice, nu atât de neconvenţionale încât să ne interzică acest lucru – dar în mod hotărât neconvenţionale – după care ar trebui să existe anumite îndoieli legate de adevărul matematic absolut al afirmaţiilor matematice idealizate. Va trebui să aruncăm cel puţin o privire spre câteva dintre ele.

Q9. Poziţia cunoscută ca *intuiţionism* nu ne permite să deducem că un calcul trebuie să se termine, la un moment dat, numai pentru că continuarea sa indefinită duce la o contradicţie; pe de altă parte, există punctele de vedere 'constructivist' şi 'finitist'. În conformitate cu acestea, n-ar trebui ca raţionamentul de tip Gödel să nu fie pus sub semnul întrebării?

In raţionamentul de tip Gödel, pe care l-am dat, am utilizat în (M) un argument de forma: 'Presupunerea că X este falsă ne conduce la o contradicţie; de aici rezultă că X este adevărată.' Aici 'X' este propoziţia '$C_k(k)$ nu se opreşte'. Acesta este un argument de tip reducere la absurd – şi, într-adevăr, argumentaţia gödeliană în întregul ei, este de acest tip. Poziţia cunoscută în matematică ca *intuiţionism* (iniţiată de matematicianul olandez L.E.J. Brouwer cam pe la 1912; conform cu Kleene 1952, şi **MN** p. 128-129) neagă posibilitatea de a face un raţionament valid utilizând reducerea la absurd. Intuiţionismul a înflorit ca reacţie la anumite tendinţe din matematică, care au apărut la sfârşitul secolului XIX şi începutul secolului XX, în conformitate cu care, existenţa unui obiect matematic poate fi *afirmată* chiar dacă nu se cunoaşte nici un mijloc de a-l construi efectiv. Uneori o folosire liberă a unui concept nebulos privind existenţa matematică poate, cu adevărat, să conducă la o contradicţie. Cel mai celebru exemplu de acest tip apare cu mulţimea paradoxală a lui Bertrand Russell „mulţimea tuturor mulţimilor care nu se conţin ca element" (dacă mulţimea Russell este un element în ea însăşi, atunci nu este; dacă nu este, atunci este! vezi şi paragraful 3.4 şi **MN** p. 114-115,

pentru mai multe detalii). Pentru a se opune tendinţei generale prin care obiectele matematice foarte liber definite pot fi luate drept existente, poziţia 'intuiţionistă' interzice considerarea validă a raţionamentului matematic care permite să fie dedusă existenţa unui obiect matematic numai din natura contradictorie a non-existenţei sale. O demonstraţie bazată pe reducere la absurd nu furnizează construcţia efectivă a obiectului în cauză.

Este afectat raţionamentul de tip Gödel de interzicerea folosirii reducerii la absurd? În realitate, nu este afectat deloc, pentru că noi am folosit reducerea la absurd invers decât am spus mai înainte, adică am ajuns la o contradicţie, din presupunerea că ceva *există* şi nu presupunând că nu există. În conformitate cu intuiţioniştii, este perfect legitim să deduci că ceva nu există din faptul că, presupunându-i existenţa, ajungi la o contradicţie. Aşa cum l-am dat, raţionamentul de tip Gödel este perfect acceptabil din punct de vedere intuiţionist (Kleene 1952).

Observaţii asemănătoare se aplică tuturor punctelor de vedere finitiste sau constructiviste de care am cunoştinţă. Discuţia care urmează întrebării Q8 arată că, chiar şi punctul de vedere de mai sus, care refuză să accepte că şirul numerelor naturale 'cu adevărat' poate fi continuat indefinit, nu ne furnizează un mijloc de a evita concluzia că, pentru a accede la adevărul matematic, nu este folosit un algoritm corect şi care poate fi cunoscut.

2.7. Câteva consideraţii matematice mai profunde

Pentru a pătrunde mai adânc implicaţiile argumentelor lui Gödel ar fi util să ne întoarcem spre ceea ce a fost de fapt punctul de pornire. La sfârşitul secolului trecut şi începutul secolului 20, cei care se ocupau de fundamentele matematicii au fost confruntaţi cu dificultăţi serioase.

La sfârşitul secolului 19, datorită mai ales contribuţiei originale şi profunde a matematicianului Georg Cantor (a cărui metodă a 'diagonalizării', am discutat-o mai înainte), matematicienii au găsit metode puternice pentru a stabili unele dintre rezultatele lor cele mai adânci, bazându-şi argumentele pe proprietăţile mulţimilor infinite. În acelaşi timp, au apărut dificultăţi legate tocmai de mulţimile infinite, atunci când acestea sunt folosite într-un mod prea liber. În particular, paradoxul lui Russell (pe care l-am amintit în răspunsul la întrebarea Q9 şi în paragraful 3.4 şi

care a fost observat chiar de Cantor) pune în evidenţă câteva obstacole în utilizarea mulţimilor infinite într-un mod prea dezinvolt. Totuşi, s-au putut obţine rezultate interesante, cu condiţia ca modul de raţionament aplicat să fie urmărit cu suficientă atenţie. Se punea problema stabilirii condiţiilor în care un anumit raţionament se făcea cu 'suficientă atenţie'.

Marele matematician David Hilbert a fost una din figurile cele mai importante în mişcarea care avea ca scop 'securizarea' raţionamentului matematic. Această mişcare a fost denumită formalism.

În conformitate cu doctrina formalismului, toate formele permise ale raţionamentului matematic dintr-un domeniu specific, care includ şi apelează la raţionamente asupra mulţimilor infinite, trebuie să poată fi descrise univoc. Un astfel de sistem de reguli şi propoziţii matematice se numeşte sistem formal. Odată ce regulile unui sistem formal $\mathbb{F}$ au fost fixate, mai departe se verifică mecanic dacă regulile, în număr necesar finit[6], au fost sau nu bine aplicate. Desigur, regulile trebuie privite ca forme valide ale raţionamentului matematic, astfel încât orice rezultat, care poate fi obţinut aplicându-le, să poată fi garantat ca fiind de fapt *adevărat*. Totuşi, unele dintre aceste reguli presupun utilizarea mulţimilor infinite şi aici, anumite intuiţii matematice, ca cele ce privesc stabilirea formelor de raţionament care sunt legitime şi care nu, ar putea să nu fie de absolută încredere. Îndoielile în acest context ar putea fi îndreptăţite dacă se ţine seama de contradicţiile care apar în caz că se permite o utilizare nerestrictivă a mulţimilor infinite, precum „mulţimea tuturor mulţimilor care nu se conţin ca element" ce intervine în paradoxul lui Russell. Regulile din $\mathbb{F}$ trebuie să împiedice utilizarea mulţimii lui Russell, dar până la ce nivel? Să se interzică complet utilizarea mulţimilor infinite; ar fi mult prea restrictiv (de exemplu, spaţiul Euclidian obişnuit conţine o infinitate de puncte, mulţimea numerelor naturale este şi ea infinită); mai mult, s-a constatat că există numeroase sisteme formale particulare (satisfăcătoare, în sensul că nu permit, spre exemplu, definirea unei mulţimi ca cea a lui Russell), cu ajutorul cărora se pot deduce cele mai multe din

---

[6] Unele sisteme formale sunt prezentate ca având un număr *infinit* de axiome - descris în termenii unor structuri cunoscute drept „scheme axiomatice" dar, pentru a denumi sistem formal în sensul în care eu îl folosesc aici, sistemul formal trebuie să poată fi exprimat în termeni finiţi, un sistem axiomatic infinit putând fi generat de o mulţime finită de reguli de calcul. Chiar aşa se întâmplă în cazul sistemelor formale utilizate în demonstraţiile matematice ca de exemplu, sistemul formal bine-cunoscut Zermelo - Fraenkel $\mathbb{ZF}$ care descrie teoria convenţională a mulţimilor.

rezultatele matematice dorite. Cum se poate stabili care din aceste sisteme formale sunt de încredere şi care nu?

Să ne fixăm atenţia asupra unui astfel de sistem formal $\mathbb{F}$ şi să folosim notaţia ADEVĂRAT şi FALS, pentru propoziţiile ce pot fi obţinute cu ajutorul regulilor din $\mathbb{F}$ şi, respectiv, ale căror negaţii se pot obţine cu ajutorul regulilor din $\mathbb{F}$. Orice propoziţie ce poate fi formulată în interiorul sistemului $\mathbb{F}$, dar care nu este nici ADEVĂRATĂ, nici FALSĂ, în sensul de mai sus, va fi NEDECIDABILĂ.

Unii cercetători vor obiecta, considerând că, atât timp cât mulţimile infinite pot fi 'fără sens', problema adevărului sau falsităţii nu s-ar pune în ceea ce le priveşte (acest lucru s-ar aplica cel puţin unor tipuri anume de mulţimi infinite, dacă nu tuturor). Conform acestui punct de vedere, nu ar conta care propoziţii despre (anumite) mulţimi infinite s-ar constata că sunt ADEVĂRATE şi care FALSE, cu condiţia ca nici o propoziţie să nu fie *în acelaşi timp* şi ADEVĂRATĂ şi FALSĂ. Această presupunere revine la a spune că sistemul $\mathbb{F}$ este *consistent*[7] Pentru aceştia – adevăraţii *formalişti* — singurele probleme importante privind un sistem formal ar fi: (a) dacă sistemul este sau nu *consistent*,

(b) dacă este sau nu *complet*.

Despre sistemul $\mathbb{F}$ se zice că este complet dacă orice propoziţie matematică, corect formulată în $\mathbb{F}$, este întotdeauna sau ADEVĂRATĂ sau FALSĂ (adică $\mathbb{F}$ nu conţine nici o propoziţie NEDECIDABILĂ).

Pentru un formalist în sens strict, întrebarea dacă o propoziţie despre mulţimi infinite este sau nu *actual adevărată*, nu este în mod necesar semnificativă, în mod cert nefiind considerată relevantă pentru procedurile matematicii formale. Astfel, în loc să se preocupe de adevărul matematic absolut al propoziţiilor de acest tip, se cere demonstrarea consistenţei şi completitudinii sistemelor formale. Ce fel de reguli matematice sunt permise pentru acest tip de demonstraţii? Aceste reguli trebuie să fie de absolută încredere şi să nu facă uz de nici un fel de raţionament dubios cu mulţimi infinite insuficient precizate (precum cea a lui Russell). S-a sperat că ar putea exista procedee logice în interiorul unor sisteme formale comparativ simple şi evident corecte (ca sistemul relativ elementar cunoscut drept aritmetica lui Peano) care să fie suficiente pentru a demonstra consistenţa altor sisteme formale sofisticate — să zicem $\mathbb{F}$ — care ar putea permite raţionamente

[7] In limba română, se foloseşte uneori şi termenul *necontradictoriu* cu aceeaşi semnificaţie (n.t.)

formale asupra unor mulţimi infinite 'foarte mari', dar a căror consistenţă să nu fie deloc evidentă. Dacă este acceptată filosofia formalistă, atunci, o astfel de demonstraţie a consistenţei lui $\mathbb{F}$ ar furniza şi justificarea utilizării mijloacelor de raţionament permise de $\mathbb{F}$. Atunci, demonstraţii ale teoremelor de matematică s-ar putea face folosind mulţimile infinite într-un mod consistent, dispensându-ne de analiza semnificaţiei actuale a acestor mulţimi. Mai mult, dacă despre un astfel de sistem $\mathbb{F}$ s-ar putea arăta că este complet, atunci în mod îndreptăţit s-ar putea crede că $\mathbb{F}$ concentrează, de fapt, *toate* procedurile matematice permise; astfel într-un anumit sens, $\mathbb{F}$ ar putea fi considerat ca *fiind* formularea completă a matematicii domeniului respectiv. În aceste condiţii, în 1930, Gödel a produs rezultatul său revoluţionar exploziv (publicat în 1931), care a arătat, în cele din urmă, că visul formaliştilor era de neatins. Gödel a demonstrat că nu există un sistem formal $\mathbb{F}$ care să fie totodată şi consistent (într-un anumit sens 'tare', pe care-l vom descrie în secţiunea următoare) şi complet când despre $\mathbb{F}$ se presupune că este suficient de puternic pentru a conţine o formulare a propoziţiilor din aritmetica obişnuită şi logica standard. Astfel, teorema lui Gödel s-ar aplica unor sisteme $\mathbb{F}$ pentru care propoziţiile aritmeticii, ca de exemplu teorema lui Lagrange şi conjectura lui Goldbach, aşa cum au fost descrise în paragraful 2.3, pot fi formulate ca propoziţii în sistemul formal.

În discuţiile care vor urma mă voi concentra asupra sistemelor formale suficient de extinse pentru ca operaţiile aritmetice necesare pentru formularea teoremei lui Gödel să fie conţinute în sistemul formal (şi ca operaţiile oricărei maşini Turing să fie conţinute în el, dacă va fi necesar). Deci, atunci când ne vom referi la un anumit sistem formal $\mathbb{F}$ vom presupune că el este suficient de extins (în sensul de mai înainte). Acest lucru nu restrânge în nici un fel discuţia (cu toate acestea, pentru claritate, voi adăuga cuvintele 'suficient de extins', în legătură cu un sistem formal în discuţie).

2.8. Condiţia de ω-consistenţă

Forma cea mai cunoscută a teoremei lui Gödel spune că, un sistem formal $\mathbb{F}$ suficient de extins nu poate fi în acelaşi timp şi complet şi consistent. Aceasta nu-i chiar faimoasa 'teoremă de incompletitudine' care a fost comunicată pentru

prima dată la întâlnirea de la Königsberg la care ne-am referit în paragrafele 2.1 şi 2.7, ci este o versiune ceva mai tare, obţinută mai târziu de logicianul american J. Barclay Rosser (1936).

Versiunea pe care Gödel a anunţat-o iniţial era echivalentă cu a arăta că $\mathbb{F}$ nu poate fi în acelaşi timp şi complet şi *ω-consistent*. Condiţia de ω-consistenţă este cu ceva mai tare decât condiţia de consistenţă. Pentru a explica ce înseamnă acest lucru avem nevoie de câteva notaţii noi. Dintre notaţiile utilizate într-un sistem formal fac parte simbolurile logice. Trebuie să existe un simbol pentru *negaţie*, acesta este de obicei ~. Astfel, dacă Q este o propoziţie în $\mathbb{F}$ atunci $\sim Q$ notează propoziţia 'non-Q' Trebuie să existe, de asemeni, un simbol care spune 'oricare ar fi [numărul natural]', simbol care se numeşte *cuantificator universal* şi se notează de obicei cu $\forall$ Dacă $P(n)$ este o propoziţie ce depinde de numărul natural n (în acest caz P este ceea ce se numeşte o funcţie propoziţională), atunci şirul de simboluri $\forall n[P(n)]$ desemnează propoziţia „pentru orice număr natural n este adevărată afirmaţia $P(n)$" Un exemplu particular de astfel de propoziţie $P(n)$ ar fi: „Numărul n poate fi exprimat ca o sumă de trei pătrate" şi atunci $\forall n[P(n)]$ înseamnă: „orice număr natural este suma a trei pătrate", care, în acest caz este falsă (propoziţia este adevărată dacă în loc de 'trei' scriem 'patru'). Putem combina astfel de simboluri în multe feluri; în particular, şirul de simboluri $\sim\forall n[P(n)]$ exprimă negaţia faptului că $P(n)$ ar fi adevărată pentru orice număr natural n

Condiţia de ω-consistenţă spune că dacă $\sim\forall n[P(n)]$ poate fi demonstrată prin metodele din $\mathbb{F}$ atunci nu toate propoziţiile $P(0)$, $P(1)$, $P(2)$, $P(3)$, $P(4)$,... sunt deductibile în $\mathbb{F}$. De aici rezultă,că, dacă $\mathbb{F}$ nu ar fi ω-consistent, am avea situaţia anormală în care, pentru anumiţi P, fiecare dintre propoziţiile $P(0)$, $P(1)$, $P(2)$, $P(3)$, $P(4)$,... ar putea fi demonstrată; totodată afirmaţia că nu toate aceste propoziţii sunt adevărate ar fi şi ea demonstrată. Evident, nici un sistem formal demn de încredere nu poate admite un asemenea lucru. Dacă sistemul formal $\mathbb{F}$ este *corect*[8], atunci el este cu siguranţă ω-consistent.

În această carte vom nota cu $G(\mathbb{F})$ şi cu $\Omega(\mathbb{F})$ propoziţiile:

$G(\mathbb{F})$ „Sistemul formal $\mathbb{F}$ este consistent" şi respectiv,

[8] Un sistem formal $\mathbb{F}$ este corect (*sound*) dacă orice teoremă a sistemului este validă (n.t.)

$\Omega(\mathbb{F})$ „Sistemul formal $\mathbb{F}$ este ω-consistent"

De fapt (presupunând că $\mathbb{F}$ este suficient de extins), $G(\mathbb{F})$ şi $\Omega(\mathbb{F})$ sunt propoziţii care pot fi formulate, la rândul lor, în termenii operaţiilor din $\mathbb{F}$. Faimoasa teoremă de incompletitudine a lui Gödel ne spune că $G(\mathbb{F})$ nu este o teoremă a lui $\mathbb{F}$ (adică nu poate fi demonstrată cu metode permise de $\mathbb{F}$) şi, la fel, nici $\Omega(\mathbb{F})$ Şi aceasta cu condiţia ca $\mathbb{F}$ să fie cu adevărat consistent. Versiunea oarecum mai tare a teoremei lui Gödel, obţinută mai târziu de Rosser, spune că dacă $\mathbb{F}$ este un sistem consistent, atunci nici $\sim G(\mathbb{F})$ nu este o teoremă în $\mathbb{F}$ În restul acestui capitol voi încerca să prezint argumentaţia mea în termenii lui $G(\mathbb{F})$, mai familiari şi nu ai lui $\Omega(\mathbb{F})$, deşi, pentru cea mai mare parte a discuţiei, ambele pot fi la fel de bune (Pentru anumite argumente mai explicite, din capitolul 3, voi folosi uneori notaţia $G(\mathbb{F})$ pentru propoziţia $C_k(k)$ nu se opreşte' — conform cu paragraful 2.5 — ceea ce, de fapt, nu este un abuz de notaţie).

În continuare, în discuţiile care urmează, nu voi trage o linie clară între consistenţă şi ω-consistenţă, dar versiunea teoremei lui Gödel, pe care am prezentat-o în paragraful 2.5, este, în esenţă, cea care susţine că, dacă sistemul $\mathbb{F}$ este consistent el nu poate fi complet deoarece este incapabil să susţină că $G(\mathbb{F})$ este o teoremă. Nu intenţionez să demonstrez acest lucru aici (dar vezi Kleene 1952). De fapt, pentru ca această formă a raţionamentului lui Gödel să fie reductibilă la argumentaţia dată aici de mine, se cere ceva mai mult sistemului $\mathbb{F}$ decât „să conţină aritmetica şi logica obişnuită"

Mai avem nevoie ca sistemul $\mathbb{F}$ să fie suficient de larg încât să includă acţiunile unei maşini Turing. Astfel, mulţimea propoziţiilor care pot fi corect formulate utilizând simbolurile unui sistem $\mathbb{F}$ trebuie să includă propoziţii de forma:

„O maşină Turing de un anumit tip, atunci când acţionează asupra numărului natural n, produce numărul natural p "

De fapt, aceasta este o teoremă (conform cu Kleene 1952, capitolele 11 şi 13) care are loc în mod automat dacă $\mathbb{F}$ include, în plus, faţă de operaţiile aritmetice obişnuite şi operaţia (numită μ):

„Găseşte cel mai mic număr natural având o anumită proprietate prestabilită"

Să reamintim că, în exemplul nostru de calcul (A), procedura noastră găsea cel mai mic număr care nu este suma a trei pătrate. În general, algoritmilor de calcul trebuie să li se permită aşa ceva.

Într-adevăr, această ipoteză ne conduce la posibilitatea de a întâlni calcule care nu se termină, ca cele din (**B**), unde am încercat să găsim cel mai mic număr care nu este suma a patru pătrate perfecte, când de fapt un astfel de număr nu există.

2.9. Sistemele formale şi demonstraţia algoritmică

În paragraful 2.5 ne-am referit, în legătură cu raţionamentul Gödel-Turing, mai cu seamă la 'calcule' şi, deloc, la 'sisteme formale'. Dar între cele două concepte există o legătură foarte strânsă. Una din proprietăţile esenţiale ale unui sistem formal este că trebuie să existe un procedeu algoritmic F (computaţional, care să permită calculul), care să poată verifica dacă regulile din $\mathbb{F}$ au fost sau nu corect aplicate.

Dacă o propoziţie este ADEVĂRATĂ, în conformitate cu regulile lui $\mathbb{F}$, atunci algoritmul de calcul F va confirma acest lucru (iar F nu face altceva decât că trece prin toate variantele posibile de şiruri de simboluri din alfabetul sistemului $\mathbb{F}$, pentru a se opri cu succes când propoziţia dorită P este întâlnită ca şir final, toţi paşii succesivi respectând regulile sistemului $\mathbb{F}$).

Reciproc, dacă E este o procedură de calcul, anume construită pentru a stabili adevărul unor propoziţii matematice, atunci putem găsi un sistem formal $\mathbb{E}$ care să confirme ca ADEVĂRATE toate adevărurile care se obţin cu ajutorul procedurii E. Trebuie să observăm că sistemul formal include operaţiile logice standard, în vreme ce procedura E s-ar putea să nu fie suficient de extinsă pentru a le incorpora direct. Dacă procedura E dată, nu încorporează în sine operaţiile logice elementare, acestea vor fi adăugate atunci când din E se construieşte sistemul formal $\mathbb{E}$, astfel că propoziţiile ADEVĂRATE în $\mathbb{E}$ nu vor fi numai propoziţiile ce se pot obţine direct prin procedura E, ci şi acele propoziţii care sunt consecinţe ale logicii elementare. În acest caz, sistemul $\mathbb{E}$ nu va fi strict echivalent cu procedura E ci va fi întrucâtva mai puternic.

Operaţiile logice considerate arată cam aşa: „dacă $P\&Q$ atunci P „dacă P şi $P \Rightarrow Q$ atunci Q„; „dacă $\forall x[P(x)]$ atunci $P(n)$„; „dacă $\sim\forall x[P(x)]$ atunci $\exists x[\sim P(x)]$ etc. Aici simbolurile $\&$, $\Rightarrow$, $\forall$, $\exists$, $\sim$ sunt respectiv operatorii 'şi', 'implicaţie', 'oricare ar fi [numărul natural]', 'există [numărul

natural]' 'negaţie'. Dar mai pot apare şi alte câteva simboluri.

Pentru a construi sistemul formal $\mathbb{E}$ pornind de la procedura E, va trebui să considerăm un sistem formal consistent de bază $\mathbb{L}$, exprimând regulile primitive ale inferenţei logice, aşa cum apar ele în sistemul cunoscut drept calculul predicatelor (Kleene 1952) şi să-l construim pe $\mathbb{E}$ adăugând lui $\mathbb{L}$, procedura de calcul E, în forma unor axiome adiţionale şi reguli pentru $\mathbb{L}$ şi, prin aceasta, considerând orice propoziţie P ca ADEVĂRATĂ dacă poate fi obţinută prin procedura E. Aceasta poate să nu fie simplu de făcut în practică. Dacă E este doar specificarea unei maşini Turing, vom adăuga sistemului $\mathbb{L}$ toate notaţiile necesare definirii unei maşini Turing, ca parte a alfabetului şi a regulilor din $\mathbb{L}$, înainte de a adăuga procedura E propriu-zisă drept axiomă adiţională (vezi şi paragraful 2.8; pentru detalii Kleene 1952).

Pentru scopul pe care-l urmărim, de fapt, nu este important dacă sistemul formal $\mathbb{E}$, construit aşa cum am arătat, poate conţine şi alte propoziţii ADEVĂRATE, altele decât cele ce se obţin direct din E (regulile logice primitive din $\mathbb{L}$ nu sunt, în mod necesar, reprezentate ca parte a procedurii E). Ne-am ocupat în paragraful 2.5 de presupusul algoritmul A, despre care am spus că ar conţine toate procedurile (cunoscute sau care pot fi cunoscute) pe care le-ar putea folosi matematicienii pentru a arăta că anumite calcule nu se opresc. Orice astfel de algoritm trebuie cu siguranţă să conţină, printre altele, toate operaţiile de inferenţă logică simplă. Pentru discuţia care urmează vom presupune că A în adevăr încorporează aceste elemente.

Pentru argumentaţia noastră, algoritmii adică procedurile de calcul şi sistemele formale pot fi considerate echivalente, ca mijloace care ne conduc la adevărurile matematice. De aceea, deşi argumentaţia din paragraful 2.5 a fost numai în termeni de calculabilitate, ea este relevantă pentru sistemele formale, în general. Să ne amintim că argumentaţia se referea la o înşiruire (listare) a tuturor calculelor (acţiunilor unei maşini Turing) $C_q(n)$ Pentru a o transpune în detalii pentru un sistem formal $\mathbb{F}$, trebuie ca sistemul $\mathbb{F}$ să fie suficient de larg pentru a încorpora acţiunile tuturor maşinilor Turing. Procedura algoritmică A utilizată pentru a stabili dacă anumite calcule nu se termină, poate fi acum încorporată printre regulile lui $\mathbb{F}$ astfel încât toate calculele pentru care proprietatea de a nu se termina este ADEVĂRATĂ în $\mathbb{F}$ să coincidă cu acelea determinate de A că nu se termină.

Ce legătură există între raţionamentul lui Gödel, prima dată prezentat la Königsberg şi cel pe care l-am expus în paragraful 2.5 ? Nu am să dau aici detalii dar voi marca paşii principali. Procedura algoritmică A juca acolo rolul sistemului formal $\mathbb{F}$ din teorema lui Gödel.

algoritmul A $\Longleftrightarrow$ regulile din $\mathbb{F}$

Propoziţia particulară „$C_k(k)$ nu se opreşte", obţinută în paragraful 2.5, care este inaccesibilă prin procedura A, dar care poate fi apreciată ca adevărată, atâta vreme cât presupunem că A este corectă, joacă aici rolul propoziţiei $G(\mathbb{F})$, propoziţie pe care Gödel a prezentat-o la Königsberg, care susţine de fapt că $\mathbb{F}$ este un sistem formal consistent:

declaraţia $C_k(k)$ nu se opreşte $\Longleftrightarrow$ afirmaţia că $\mathbb{F}$ este consistent.

Acest lucru ne poate ajuta să înţelegem cum încrederea în corectitudinea unei proceduri precum A, ne poate conduce la o altă procedură aflându-se dincolo de scopul procedurii iniţiale, despre care va trebui să credem că este şi ea corectă. Pentru că dacă credem că procedurile unui anumit sistem formal $\mathbb{F}$ sunt corecte, adică ele ne permit să deducem numai adevăruri matematice (şi nu falsuri), astfel încât, dacă despre o propoziţie A se deduce că este ADEVĂRATĂ , atunci ea trebuie de fapt să fie adevărată — atunci trebuie să credem de asemenea că $\mathbb{F}$ este ω-consistent. Dacă ADEVĂRAT implică 'adevărat' şi FALS implică 'fals' — aşa cum se întâmplă în orice sistem formal corect $\mathbb{F}$, atunci cu siguranţă:

nu toate propoziţiile $P(0)$, $P(1)$, $P(2)$, $P(3)$, $P(4)$,... pot fi ADEVĂRATE dacă este FALS că $P(n)$ are loc pentru toate numerele naturale n ,

ceea ce, până la urmă este proprietatea de ω-consistenţă. Încrederea în corectitudinea lui $\mathbb{F}$ îndreptăţeşte atât încrederea în ω-consistenţa lui, cât şi încrederea în consistenţa lui. Pentru că, dacă 'ADEVĂRAT' implică 'adevărat' şi 'FALS' implică 'fals', atunci cu siguranţă:

nici un P nu poate fi, în acelaşi timp, şi ADEVĂRAT şi FALS ,

adică tocmai ceea ce consistenţa sistemului stabileşte. De fapt, pentru multe sisteme, distincţia dintre consistenţă şi ω-consistenţă dispare. În cele ce urmează nu voi mai insista asupra distincţiei dintre cele două tipuri de consistenţă, ci voi vorbi despre *consistenţă*, pentru ca lucrurile să fie mai simple. Pentru Gödel şi Rosser consistenţa unui sistem formal (suficient de extins) este ceva ce se află în

afara puterii sistemului formal însuşi (de a o stabili). Prima versiune a teoremei lui Gödel (Königsberg), depinde de ω-consistenţă, dar forma ulterioară , mai bine cunoscută, se referă doar la consistenţa obişnuită.

Contribuţia raţionamentului lui Gödel, din punctul nostru de vedere, constă în faptul că arată cum se poate trece dincolo de orice mulţime de reguli de calcul, despre care credem că sunt corecte, pentru a găsi o altă regulă care nu este conţinută printre cele dintâi şi despre care trebuie să credem că este de asemeni corectă, adică regula care susţine consistenţa regulilor iniţiale. Esenţial pentru noi este:

Încrederea în '*corectitudine*' implică încrederea în '*consistenţă*'.

Nu avem nici un drept să utilizăm regulile unui sistem formal $\mathbb{F}$ şi să credem că rezultatele ce pot fi deduse ar fi de fapt adevărate, decât dacă credem de asemenea în consistenţa sistemului formal. (De exemplu, dacă $\mathbb{F}$ ar fi inconsistent atunci am putea deduce ca ADEVĂRATĂ propoziţia „1 = 2", care, cu siguranţă, nu este adevărată.) Astfel, dacă credem că de fapt facem matematică atunci când utilizăm un sistem formal $\mathbb{F}$ atunci trebuie să fim pregătiţi să acceptăm raţionamentul care trece dincolo de limitările sistemului $\mathbb{F}$, oricare ar fi acesta.

2.10. Alte obiecţii tehnice posibile relativ la $\mathcal{G}$

Să examinăm în continuare alte obiecţii care mi s-au adus în timp, relativ la modul în care am utilizat raţionamentul Gödel-Turing. Multe dintre obiecţii sunt legate unele de altele, dar cred că este util să le prezint separat.

Q10. Este adevărul matematic ceva absolut? Tocmai am văzut că există diferite puncte de vedere asupra adevărului absolut al propoziţiilor care privesc mulţimile infinite. Putem avea încredere în argumente care depind de un concept vag ca cel al 'adevărului matematic', în contrast, să zicem, cu conceptul clar definit de ADEVĂR **formal?**

În cazul unui sistem formal $\mathbb{F}$ care este legat de teoria generală a mulţimilor, se poate într-adevăr să nu fie întotdeauna limpede dacă există vreun sens absolut în care o propoziţie despre mulţimi este 'adevărată' sau 'falsă' caz în care, chiar

conceptul de 'corectitudine' pentru un sistem formal precum $\mathbb{F}$, poate fi pus în discuţie. Un exemplu celebru care ne prezintă într-un mod mai familiar acest tip de problemă este conţinut într-un rezultat demonstrat de Gödel (1940) şi Cohen (1966). Ei au arătat că propoziţiile matematice numite *axioma alegerii* şi *ipoteza continuului* (ale lui Cantor) sunt independente de axiomele Zermelo-Fraenkel ale teoriei mulţimilor – un sistem formal standard pe care-l vom nota aici cu $\mathbb{ZF}$. (Axioma alegerii spune că, oricare ar fi o colecţie de mulţimi nevide, există o altă mulţime, care cu siguranţă conţine câte un element din fiecare membru al colecţiei [1]. Ipoteza continuului, a lui Cantor, spune că numărul submulţimilor mulţimii numerelor naturale – care este acelaşi cu numărul numerelor reale – este următorul infinit; 'mai mare', după infinitul numărului de numere naturale [2]. Nu este necesar ca în acest moment cititorul să aprecieze semnificaţia acestor axiome. Nici pentru mine nu-i necesar să intru în detalii privind axiomele şi regulile de procedură din $\mathbb{ZF}$.)

Unii matematicieni vor susţine că $\mathbb{ZF}$ conţine tot arsenalul necesar din matematica obişnuită. Alţii vor susţine că o argumentaţie matematică este cu siguranţă una care poate, în principiu, să fie formulată şi demonstrată în interiorul lui $\mathbb{ZF}$ (vezi discuţia de la **Q14** pentru o evaluare a modului în care se aplică raţionamentul lui Gödel unor astfel de persoane). Aceşti matematicieni vor susţine că propoziţiile matematice care sunt ADEVĂRATE, FALSE, sau NEDECIDABILE în conformitate cu $\mathbb{ZF}$ sunt, cu siguranţă, propoziţiile care, în principiu, pot fi stabilite matematic ca fiind adevărate, stabilite matematic ca fiind false, respectiv matematic indecidabile. Pentru astfel de persoane, axioma alegerii şi ipoteza continuului vor fi matematic indecidabile (aşa cum, spun aceştia, ne arată rezultatul lui Gödel-Cohen) şi ei ar putea foarte bine să argumenteze că adevărul sau falsitatea acestor afirmaţii este o chestiune pur convenţională.

De fapt, afectează aceste aparente incertitudini, cum ar fi natura absolută a adevărului matematic, deducţiile noastre rezultate din raţionamentul Gödel-Turing? Chiar, deloc; deoarece ne ocupăm aici de o clasă de probleme matematice de natură mult mai limitată decât problemele care, ca axioma alegerii sau ipoteza continuului, intervin în cazul mulţimilor infinite non-constructibile. Ne vom ocupa aici doar de propoziţii de forma: „cutare şi cutare calcul nu se termină niciodată" şi unde calculul despre care-i vorba poate fi specificat cu precizie în termenii acţiunilor unei maşini Turing. Astfel de propoziţii sunt cunoscute de logicieni ca propoziţii-π_1

(sau încă mai corect, ca propoziţii-π_1^0).

Pentru orice sistem formal $\mathbb{F}$ propoziţia $G(\mathbb{F})$ este o propoziţie-π_1 (dar nu şi $\Omega(\mathbb{F})$) (vezi paragraful 2.8). S-ar părea că ar fi prea puţin rezonabilă îndoiala că natura adevărată/falsă a oricărei propoziţii-π_1 ar fi una absolută, independent ce poziţie am adopta în chestiunile care privesc mulţimile infinite non-constructibile – ca cele care implică axioma alegerii şi ipoteza continuului (pe de altă parte, aşa cum vom vedea imediat, tipul de raţionament despre care se acceptă că furnizează demonstraţii convingătoare ale propoziţiilor-π_1 , ar putea, cu adevărat, să depindă de poziţia cu privire la non-constructivitatea mulţimilor infinite; conform cu **Q11**). Pare clar că, în afara acelor poziţii extreme, luate de anumiţi intuiţionişti (vezi răspunsul la **Q9**), singura poziţie rezonabilă cu privire la natura absolută a adevărului unor astfel de propoziţii pare să fie aceea că anumite calcule care se termină pot să dureze atât de neobişnuit de mult încât ele să nu poată fi practic încheiate, să zicem în toată istoria universului; sau în cazul în care calculul în sine necesită atât de multe simboluri pentru a fi specificat (deşi sunt în număr finit) încât nu ar putea fi niciodată scris în întregime. Dar astfel de situaţii au fost deja complet analizate în discuţia relativă la **Q8** şi am văzut acolo că concluzia noastră esenţială $\mathcal{G}$ a rămas intactă. Să ne mai amintim şi că, legat de discutarea lui **Q9**, poziţia intuiţionistă nu se sustrage nici ea concluziei $\mathcal{G}$.

Ar mai fi ceva, foarte limitatul concept de adevăr matematic, de care am nevoie pentru raţionamentul Gödel-Turing, nu este cu nimic mai puţin bine definit decât conceptele ADEVĂRAT, FALS şi NEDECIDABIL din orice sistem formal $\mathbb{F}$ Să reamintim, din cele spuse anterior (paragraful 2.9), că există un *algoritm* F care este echivalent cu $\mathbb{F}$ Dacă F este prezentat printr-o propoziţie P (formulabilă/asertabilă în limbajul lui $\mathbb{F}$) atunci acest algoritm ajunge cu succes la oprire exact când propoziţia P este demonstrabilă în conformitate cu regulile din $\mathbb{F}$, adică atunci când P este ADEVĂRATĂ. La fel, propoziţia P este FALSĂ atunci când F ajunge cu succes la oprire, după ce apare $\sim P$ de asemenea, P va fi NEDECIDABILĂ exact atunci când nici unul din aceste calcule nu se termină.

Chestiunea dacă propoziţia matematică P este ADEVĂRATĂ, FALSĂ, sau NEDECIDABILĂ, are cu siguranţă acelaşi caracter ca şi adevărul actual al terminării sau neterminării calculelor adică ca acela al adevărului sau falsităţii unor propoziţii-π_1 anume exact ceea ce este necesar pentru argumentaţia noastră de tip Gödel-Turing.

Q11. Există anumite propoziţii-π_1 care pot fi demonstrate utilizând teoria mulţimilor infinite, dar pentru care nu se cunosc demonstraţii utilizând metode standard „finite". Oare nu înseamnă acest lucru că modul în care matematicienii decid, chiar în cazul unor astfel de întrebări bine-definite, este în realitate subiectiv? Matematicieni diferiţi, cu convingeri diferite cu privire la teoria mulţimilor, ar putea avea criterii neechivalente pentru evaluarea veridicităţii propoziţiilor-π_1

Acesta ar putea fi un punct important legat de deducţiile mele din raţionamentul Gödel(-Turing) şi probabil că nu i-am dat destulă importanţă în prezentarea din **MN**. În mod surprinzător, **Q11** nu pare să fi fost o obiecţie care să fi neliniştit pe altcineva decât pe mine, în orice caz, nimeni nu mi-a adus-o la cunoştinţă. Atât aici cât şi în cartea citată (paginile 450-452) am transpus rezultatul Gödel-Turing în sensul că 'matematicienii' sau 'comunitatea matematică' sunt capabili să ajungă la anumite concluzii pe baza raţionamentului şi a intuiţiei. Avantajul de a vorbi în aceşti termeni, fără a mă referi la un anumit individ, care poate ajunge la concluzii utilizându-şi propria raţiune sau intuiţie, este că ne permite să evităm anumite obiecţii frecvent aduse formulării lui Lucas (1961). Mulţi [3] au obiectat că, de exemplu, chiar 'Lucas însuşi' s-ar putea să nu-şi poată cunoaşte propriul algoritm (astfel de persoane au adus acelaşi tip de obiecţii prezentării mele [4], deşi eu nu mi-am prezentat argumentaţia într-un fel 'personal'). Avantajul de a ne referi la raţiunea şi intuiţia 'matematicienilor', sau a 'comunităţii matematice' este că acest lucru ne permite să scăpăm de impresia că diferite individualităţi pot percepe adevărul matematic în moduri diferite, fiecare în acord cu propriul algoritm care nu poate fi cunoscut. Este mult mai greu de acceptat că capacitatea de înţelegere generală din 'comunitatea matematicienilor' ca întreg, ar putea fi rezultatul unui algoritm necunoscut, aşa cum s-ar putea crede în cazul particular al unui individ. Problema pe care o ridică întrebarea **Q11** este că această înţelegere împărtăşită de matematicieni ar putea să nu fie chiar atât de universală şi impersonală precum am prezentat-o eu.

Desigur, este adevărat că există propoziţii ca cele la care se referă **Q11**. Astfel, există propoziţii-π_1, ale căror demonstraţii cunoscute depind de utilizarea corespunzătoare a teoriei mulţimilor infinite. O astfel de propoziţie-π_1 poate să apară din codificarea aritmetică a unei propoziţii de tipul 'axiomele lui $\mathbb{F}$ sunt

consistente', unde sistemul formal $\mathbb{F}$ implică manipularea unor mulţimi infinite a căror existenţă ar putea fi controversată. Un matematician care crede în existenţa actuală a unor mulţimi neconstructive S va ajunge la concluzia că $\mathbb{F}$ este de fapt consistent, în vreme ce un altul, care nu crede în S, nu va avea nici o încredere în consistenţa lui $\mathbb{F}$. Deci, chiar restrângându-ne atenţia la chestiuni bine definite privind oprirea sau vreo altă acţiune a unei maşini Turing (adică la adevărul sau falsitatea propoziţiilor-π_1) tot nu putem ignora problema subiectivităţii credinţelor, să spunem în existenţa unei mulţimi infinite neconstructive S. Dacă feluriţi matematicieni folosesc 'algoritmi personali' neechivalenţi pentru a stabili adevărul unor propoziţii-π_1, atunci este discutabilă referirea mea la comunitatea matematicienilor.

Cred că, privind lucrurile mai îndeaproape, atitudinea mea ar putea fi discutabilă; şi cititorul care este înclinat spre o astfel de concluzie ar putea prefera o reinterpretare a lui $\mathcal{G}$:

$\mathcal{G}^*$ **Nici un matematician nu stabileşte adevărul matematic doar utilizând un algoritm despre care ştie că este corect.**

Argumentele pe care le voi folosi se pot aplica, dar cred că unele îşi pierd în bună măsură din forţă atunci când problema este prezentată în acest mod. Încă mai mult, cu versiunea $\mathcal{G}^*$, argumentaţia se îndreaptă într-o direcţie despre care cred că nu ne este de ajutor, deoarece se preocupă mai mult de mecanismul particular care guvernează acţiunile unui individ, decât de principiile care stau la baza acţiunilor oricăruia dintre noi. În acest moment mă interesează mai puţin modul în care unii matematicieni ar aborda o problemă de matematică şi mai mult să aflu ce este *universal* în mecanismele înţelegerii noastre şi în percepţia noastră matematică.

Să încercăm să vedem dacă suntem cu adevărat obligaţi să adoptăm versiunea $\mathcal{G}^*$. Sunt oare matematicienii chiar atât de subiectivi în judecata lor încât ar putea să nu fie de acord, în principiu, dacă o propoziţie-π_1 particulară, a fost sau n-a fost demonstrată ca adevărată? (Evident, demonstraţia dată pentru a stabili propoziţia-π_1 s-ar putea să fie prea lungă, sau prea complicată, pentru ca un matematician sau altul să o poată urmări conform **Q12**, de mai jos căci, cu

siguranţă, ei ar putea fi, *practic*, foarte deosebiţi. Dar nu despre asta este vorba aici. Aici ne ocupăm numai de probleme de principiu.) Pentru că, de fapt, demonstraţia matematică nu este atât de subiectivă pe cât s-ar putea crede. În ciuda faptului că diferiţi matematicieni ar putea avea diferite puncte de vedere asupra adevărului inatacabil al unor aspecte fundamentale, atunci când este vorba despre afirmarea sau infirmarea unei anumite propoziţii-π_1 clar definite, ei cad de acord, când ajung în acest punct. O propoziţie-π_1 particulară, care efectiv susţine consistenţa unui anumit sistem $\mathbb{F}$ nu va fi, în mod normal, considerată ca demonstrată acceptabil, dacă toată lumea trebuie să accepte existenţa unei anumite mulţimi infinite S, controversate. O formulare acceptabilă a ceea ce s-a demonstrat ar fi: „dacă S există, atunci sistemul $\mathbb{F}$ este consistent şi în acest caz propoziţia-π_1 este adevărată"

Totuşi, pot exista excepţii, când un matematician ar considera că o anumită mulţime neconstructivă S există în mod 'evident' — sau, cel puţin, că presupunerea existenţei sale nu ne poate duce la o contradicţie — în vreme ce un alt matematician n-ar crede deloc acest lucru. Uneori, când este vorba de aspectele *fundamentale*, se pare că matematicienii au dispute irezolvabile. În principiu acestea ar putea să ducă până acolo încât să nu poată fi convinşi să-şi comunice demonstraţiile nici măcar relativ la propoziţii-π_1. Probabil, diferiţi matematicieni au percepţii inerent diferite privind adevărul unor propoziţii legate de noconstructivitatea mulţimilor infinite. Şi este adevărat că uneori chiar *declară* că au percepţii diferite. Dar eu cred că aceste diferenţe sunt, de fapt, asemănătoare cu diferenţele dintre *aşteptările* pe care diferiţi matematicieni le-ar putea avea cu privire la adevărul propoziţiilor matematice obişnuite. Aceste aşteptări sunt, de fapt, opinii provizorii. Atâta vreme cât nu se găseşte o demonstraţie convingătoare, matematicienii se contrazic între ei cu privire la ceea ce ghicesc sau speră că-i adevărat, dar dacă unul dintre matematicieni posedă demonstraţia, acest lucru ar trebui (în principiu) să-i convingă pe ceilalţi. În ce priveşte chestiunile fundamentale, astfel de demonstraţii lipsesc. S-ar putea să nu poată fi găsite niciodată demonstraţii convingătoare. Probabil, ele nu pot fi găsite pentru că astfel de demonstraţii nu există şi astfel, există puncte de vedere diferite, la fel de îndreptăţite, cu privire la aceste chestiuni fundamentale.

În legătură cu toate acestea, totuşi trebuie să subliniem un lucru şi anume că posibilitatea ca un anumit matematician să aibă un punct de vedere eronat — şi

înţeleg prin asta un punct de vedere care permite să se tragă concluzii greşite în legătură cu o anumită propoziţie-π_1 — nu intră acum în preocupările noastre.

Posibil că matematicienii ar putea utiliza 'intuiţii greşite' şi în particular un *algoritm incorect*, dar acest lucru nu îl considerăm deocamdată important deoarece ar fi de acord cu $\mathcal{G}$. Această posibilitate va fi analizată cu atenţie şi în detaliu în paragraful 3.4. Chestiunea aici nu este dacă unii matematicieni ar putea avea puncte de vedere inconsistente, ci faptul că, un anumit punct de vedere, ar putea fi, în principiu, *mai puternic* decât un altul.

Fiecare punct de vedere ar putea fi corect în ceea ce priveşte implicaţiile sale, ca, de exemplu, adevărul propoziţiilor-π_1, dar anumite puncte de vedere ar putea să permită suporterilor lor să stabilească că anumite calcule nu se vor termina, în vreme ce, pentru punctele de vedere mai slabe, acest lucru să nu fie posibil. Şi astfel, diferiţi matematicieni ar putea avea nivele esenţial diferite de pătrundere.

Nu mă gândesc că această posibilitate ar putea conţine vreo ameninţare la adresa formulării mele iniţiale $\mathcal{G}$ Deşi ar putea exista puncte de vedere diferite cu privire la mulţimile infinite, pe care matematicienii le-ar putea susţine în mod rezonabil, care nu sunt chiar atât de numeroase probabil nu mai mult de 4-5 Singurele diferenţe importante ar fi cele de tipul axiomei alegerii (la care ne-am referit în **Q10**), pe care mulţi o consideră 'evidentă', în timp ce alţii refuză s-o accepte datorită neconstructivităţii ce o conţine. Este interesant că aceste puncte de vedere diferite cu privire la axioma alegerii însăşi nu conduc direct la vreo propoziţie-π_1, a cărei valabilitate să fie disputată. Pentru că, chiar dacă axioma alegerii este considerată 'adevărată' sau nu, această axiomă nu conduce la vreo contradicţie cu axiomele ZF standard, aşa cum rezultă din teorema Gödel-Cohen (la care ne-am referit în **Q10**). Dar ar putea să existe vreo altă axiomă disputată, pentru care să nu se cunoască nici o teoremă corespunzătoare. Dar, de obicei, când intervine acceptarea sau neacceptarea unei anumite axiome din teoria mulţimilor, să zicem axioma Q, atunci afirmaţiile matematicienilor ar trebui să ia forma: „Dacă Q este o axiomă, atunci N-ar fi loc de nici o dispută, în acest caz. Axioma alegerii pare să fie o excepţie pentru că adesea ea este presupusă fără a fi menţionată, dar, aparent, nu pune nici o problemă formulării date lui $\mathcal{G}$, cu condiţia să ne restrângem atenţia în $\mathcal{G}$, la propoziţii-π_1

$\mathcal{G}^{**}$ **Persoanele care se ocupă de matematică nu folosesc un algoritm despre care se poate şti că este corect pentru a stabili adevărul propoziţiilor-π_1 .**

Doar de acest lucru vom avea nevoie în cele ce urmează.

Oare mai există şi alte axiome disputate văzute de unii ca 'evidente' şi discutabile, de către alţii? Cred că ar fi o mare exagerare să zicem că există zece puncte de vedere esenţial diferite privind presupunerile relativ la teoria mulţimilor, care să nu fie în mod explicit menţionate ca presupuneri. Deci, să considerăm că există doar acest număr mic şi să examinăm implicaţiile. Ar însemna că există cam zece categorii diferite de matematicieni, categorii legate de tipul de raţionament privind mulţimile infinite, pe care sunt dispuşi să le accepte ca 'evident' valide. Ne putem referi la aceştia ca la matematicienii de categoria n-a, unde n ia doar câteva valori, nu mai mult de zece (să presupunem că ordonăm punctele de vedere şi cu cât n este mai mare, cu atât respectivul punct de vedere este mai puternic) În locul lui $\mathcal{G}^{**}$ vom avea:

$\mathcal{G}^{***}$ **Pentru fiecare n (n luând doar câteva valori) matematicienii de categoria n nu acceptă adevărul propoziţiilor-π_1 decât utilizând un algoritm despre care ştiu că este corect.**

Acest lucru rezultă deoarece raţionamentul Gödel Turing poate fi aplicat separat fiecărei categorii (trebuie să fie clar că raţionamentul Gödel propriu-zis nu este subiect de dispută între matematicieni aşa încât, dacă despre algoritmul presupus de categoria n s-ar şti că este corect, pentru orice matematician de categoria n, raţionamentul ar duce la o contradicţie). Astfel, ca şi în $\mathcal{G}$, nu se pune problema să existe un mare număr de algoritmi despre care să nu se ştie că sunt corecţi iar fiecare algoritm să fie specific unui anumit matematician. În loc de asta, am ajuns acum în situaţia să presupunem că ar exista doar un număr foarte mic de algoritmi despre care să nu se ştie că sunt corecţi, neechivalenţi, categorisiţi în raport cu puterea lor, care furnizează diferite şcoli de gândire.

Versiunea $\mathcal{G}^{***}$, nu diferă semnificativ de versiunile $\mathcal{G}$ şi $\mathcal{G}^{**}$ pentru discuţia care urmează şi nu voi mai face nici o distincţie între ele, denumindu-le

simplu $\mathcal{G}$

Q12. Chiar dacă matematicienii susţin că au puncte de vedere diferite în anumite chestiuni de principiu, sunt ei, în practică, chiar atât de diferiţi în ceea ce priveşte capacitatea de a urmări o demonstraţie? Şi diferă ei foarte mult în ceea ce priveşte intuiţia care să le permită să facă descoperiri în matematică?

Evident, aceste lucruri sunt adevărate, dar ele nu sunt cu adevărat relevante în ceea ce ne interesează. Pe noi nu ne interesează cu câte detalii poate un matematician să urmărească, *în practică*, o demonstraţie. Cu atât mai puţin ne preocupăm de modul în care, în practică, un anumit matematician este capabil să facă o anumită *descoperire*, sau de intuiţia şi inspiraţia care-l fac capabil de o anume descoperire. Problema este să vedem ce tip de argumente pot fi luate în considerare 'în principiu' ca valide de către matematicieni.

Să presupunem că un matematician este în posesia unei demonstraţii pentru o anumită propoziţie-π_1, (sau a unei infirmări). Dezacordul cu alţi matematicieni în legătură cu validitatea demonstraţiei se rezolvă numai dacă matematicienii au timpul, răbdarea, interesul, abilitatea şi determinarea de a urmări demonstraţia cu înţelegere şi acurateţe, chiar dacă se bazează pe un şir lung şi subtil de raţionamente. În practică, matematicienii s-ar putea să renunţe, înainte de a duce totul până la capăt. De astfel de situaţii nu ne vom ocupa aici. Aceasta pentru că certitudinea are un sens bine definit şi ceea ce este în principiu accesibil unui matematician este accesibil şi altuia (vezi consideraţiile de la **Q11**), ba mai mult, este accesibil oricărei persoane. Lanţul raţionamentului poate fi foarte lung şi conceptele implicate pot fi subtile sau obscure, dar există, fără îndoială, motive convingatoare să credem că, în modul în care o persoană înţelege un lucru, nu există nimic care, în principiu, să fie inaccesibil unei alte persoane. Acest lucru se aplică şi în cazul când se apelează la un calculator, pentru a verifica detaliile pur calculative ale unei demonstraţii.

Deşi s-ar putea să fie nerezonabil să aşteptăm ca un matematician să refacă detaliile de calcul dintr-o demonstraţie, totuşi, nu există nici o îndoială că paşii *individuali* ai demonstraţiei pot fi înţeleşi şi acceptaţi de un om care se ocupă de matematică.

Spunând toate acestea, mă gândesc doar la complexitaea unei demonstraţii matematice şi nu la chestiunile de principiu care ar putea separa un matematician

de un altul, cu privire la tipul de raţionament pe care este pregătit să-l accepte. Bineînţeles că am întâlnit matematicieni care declară că există argumente matematice care se află dincolo de competenţa lor, când spun: „Ştiu că nu voi putea niciodată înţelege cutare şi cutare lucru, oricât de tare aş încerca; tipul de raţionament mă depăşeşte" În fiecare caz în parte, când se face o astfel de declaraţie, trebuie stabilit dacă, de fapt, raţionamentul se află în afara tipului de raţionament matematic în care are încredere — aşa cum a reieşit din discuţia asupra întrebării **Q11** — sau matematicianul ar putea, dacă s-ar strădui foarte tare, să urmărească principiile care susţin argumentaţia. Cel mai adesea, chiar aşa şi este. De fapt, de cele mai multe ori, sursa disperării unui matematician provine din stilul obscur al unui articol sau, poate, dintr-o anumită limitare a capacităţii sale de lectură, nu din cauza vreunui principiu esenţial care să fie deasupra posibilităţilor sale de înţelegere. O prezentare bună, a unei chestiuni aparent obscure, poate face minuni.

Pentru a sublinia ideea pe care o susţin trebuie să mărturisesc că, adeseori, sunt prezent la seminarii de matematică unde nu pot urmări (uneori nici nu încerc să o fac!) detaliile demonstraţiilor. Am oarecum senzaţia că, dacă m-aş apuca să studiez, aş putea, în cele din urmă să urmăresc demonstraţia — deşi pentru asta ar trebui să citesc materiale suplimentare sau să primesc nişte explicaţii care să completeze detaliile lipsă din pregătirea mea de bază sau din textul seminarului. Dar eu ştiu că nu voi face de fapt nimic din toate aceste lucruri. Îmi lipsesc timpul, interesul şi entuziasmul. Astfel, accept rezultatul prezentat la seminar din tot felul de motive 'nerelevante', ca de exemplu, pentru că rezultatul 'pare plauzibil', sau pentru că vorbitorul are o reputaţie bună, sau pentru că alte persoane dintre cele prezente la seminar (despre care ştiu că sunt mult mai pătrunzătoare decât mine, în acest tip de probleme) nu au pus rezultatul sub semnul întrebării. Într-o astfel de situaţie s-ar putea să mă înşel şi rezultatul să fie, de fapt, greşit; sau poate că rezultatul este corect, dar nu rezultă din argumentele prezentate. Acestea sunt detalii irelevante pentru problema de principiu ridicată aici. Singurele excepţii ar putea surveni dacă prezentarea s-ar referi la aspecte controversate din teoria mulţimilor infinite, sau dacă ar depinde de un tip de raţionament neobişnuit, care ar putea fi îndoielnice în raport cu anumite puncte de vedere din matematică (şi care m-ar putea intriga suficient încât să mă preocup de problemă). Aceste situaţii excepţionale dar posibile sunt chiar cele de mai înainte, din **Q11**.

În legătură cu astfel de considerații asupra punctului de vedere matematic, în practică, mulți matematicieni s-ar putea să nu aibă un punct de vedere clar asupra principiilor fundamentale la care aderă. Dar, așa cum am comentat anterior, relativ la Q11, un matematician care nu are o viziune clară dacă să accepte sau nu o axiomă Q, va formula, dacă este riguros, rezultatele care îl presupun pe Q, în forma: „Presupunând axioma Q, rezultă că Evident, matematicienii, nici chiar cei mai pedanți, nu sunt totdeauna atenți la astfel de aspecte. Mai mult, este adevărat că, din când în când, fac chiar greșeli. Dar aceste greșeli, dacă sunt de fapt doar scăpări și nu au legătură cu principiile, sunt corectabile. (Așa cum am menționat anterior, posibilitatea ca ei să folosească de fapt, ca bază a deciziilor, un algoritm corect, va fi considerată în detaliu în paragrafele 3.2 și 3.4. Această posibilitate, fiind în concordanță cu $\mathcal{G}$, nu face parte din prezenta discuție.) Nu ne ocupăm aici de greșelile care sunt corectabile pentru că ele nu contribuie la ceea ce poate sau nu să fie, în principiu, obținut. Eventuala incertitudine asupra poziției unui matematician necesită o discuție mai amănunțită, așa cum urmează mai jos.

Q13. Matematicienii nu au convingeri absolut clare asupra consistenței și corectitudinii sistemelor formale pe care le folosesc — ba mai mult, nu știu sigur nici la care sistem formal aderă, de fapt. În acest caz, nu ajung convingerile lor să fie din ce în ce mai pline de umbre, pe măsură ce sistemele formale se îndepărtează de intuiția imediată și de experiență?

Într-adevăr, este un lucru rar să găsești un matematician cu opinii consistente, rigide și de neclintit, atunci când se ajunge la fundamentele unui subiect. Mai mult, pe măsură ce matematicienii câștigă experiență, punctele lor de vedere se pot modifica, relativ la ceea ce ei consideră că este inatacabil adevărat — dacă vreodată iau ceva drept inatacabil adevărat. Poate cineva să fie absolut sigur că, de exemplu, 1 diferă de 2? Dacă ar fi să vorbim despre certitudinea umană *absolută*, constatăm că nu este cu adevărat clar că un astfel de lucru există. Dar o poziție trebuie luată, totuși. Rezonabil ar fi să se aleagă un corp de principii și convingeri adevărate și de neatacat, argumentându-se pornind de aici. S-ar putea ca mulți matematicieni să nu aibă nici o opinie clară asupra a ceea ce ei consideră că este de neatacat adevărat. Într-un astfel de caz, i-aș ruga să ia totuși o poziție, chiar dacă ar fi pregătiți să o modifice ulterior. Ceea ce arată raționamentul lui Gödel este faptul că, oricare ar fi poziția adoptată, ea nu poate fi încapsulată în regulile unui sistem

formal cunoscut. Dar nu din cauza faptului că punctul de vedere s-ar modifica continuu; corpul de convingeri pe care-l circumscrie orice sistem formal $\mathbb{F}$ (suficient de extins), trebuie extins dincolo de ceea ce $\mathbb{F}$ poate atinge. Orice poziţie care include corectitudinea lui $\mathbb{F}$ printre convingerile de neatacat, trebuie să includă şi încrederea în afirmaţia lui Gödel[9] Încrederea în $G(\mathbb{F})$ nu înseamnă schimbarea poziţiei; această credinţă este conţinută implicit în punctul originar de pornire care a fost admis ca acceptându-l pe $\mathbb{F}$ — chiar dacă realizarea faptului că $G(\mathbb{F})$ trebuie să fie acceptată, poate că nu a fost evident dintru început. Bine înţeles, există întotdeauna posibilitatea să se strecoare vreo eroare cu privire la deducţiile făcute (pe baza premizelor, în orice situaţie particulară). Chiar numai posibilitatea ca cineva să facă o eroare undeva — chiar şi atunci când n-a comis-o — poate conduce la o umbrire a încrederii în concluzii. Dar acest fel de 'umbrire' nu ne priveşte de fapt. Ca şi erorile propriu zise, şi acestea sunt 'corectabile'. Mai mult, atâta vreme cât o concluzie a fost corect dedusă, cu cât este mai mult examinată, ea va fi mai convingătoare. Acest tip de neîncredere ţine ,mai mult de ceea ce simte un matematician în practica meseriei şi nu este o problemă de principiu. Astfel, suntem conduşi înapoi la discuţia de la **Q12**.

Problema este dacă există o îndoială *de principiu*, astfel încât matematicianul să considere corectitudinea unui anumit sistem formal $\mathbb{F}$ de neatacat, în vreme ce, un sistem formal mai puternic, $\mathbb{F}'$ ar putea fi corect doar 'din punct de vedere practic' Nu cred să mai fie neclarităţi, dar, oricare ar fi $\mathbb{F}$, trebuie să reamintim că el include regulile obişnuite ale operaţiilor aritmetice şi logice.

Matematicianul despre care a fost vorba mai înainte, care crede că $\mathbb{F}$ este corect, trebuie să creadă şi că $\mathbb{F}$ este consistent şi deci, că propoziţia Gödel $G(\mathbb{F})$ ataşată, este adevărată. Astfel, numai concluziile deduse din $\mathbb{F}$, singure, nu pot reprezenta totalitatea convingerilor matematice ale matematicianului, oricare ar fi $\mathbb{F}$. Dar, trebuie considerată $G(\mathbb{F})$ ca incontestabil adevărată ori de câte ori $\mathbb{F}$ este considerat incontestabil corect? Cred că nu avem de ce să ne îndoim; cu siguranţă aşa este dacă acceptăm 'în principiu' punctul de vedere pe care l-am adoptat până acum cu privire la modul în care trebuie urmărită o deducţie matematică. Singura

[9] Pentru notaţia folosită aici, vezi paragraful 2.8. Aşa cum vom vedea la sfârşitul acestui comentariu, '$\Omega(\mathbb{F})$' poate fi utilizat în locul lui $G(\mathbb{F})$

chestiune importantă priveşte detaliile codificării aserţiunii „ $\mathbb{F}$ este consistent" într-o propoziţie-π_1 aritmetică. Ideea implicată este cu siguranţă evidentă: Dacă $\mathbb{F}$ este corect, atunci cu siguranţă este consistent (căci dacă nu este consistent, atunci printre afirmaţiile sale s-ar afla ' $1 = 2$ ' şi, în acel caz, sistemul ar fi incorect). În ceea ce priveşte detaliile codificării, apare din nou distincţia dintre nivelul 'în principiu' şi nivelul 'în practică'. Nu e greu să convingi pe cineva că o astfel de codificare este posibilă, dar este cu totul altă chestiune să fii convins că orice codificare ar putea fi realizată fără greşală. Detaliile codificării pot deveni oarecum arbitrare şi ele pot să difere mult de la o expunere la alta. Ar putea să apară unele greşeli minore (sau de tipar), care, tehnic vorbind, să invalideze o anumită propoziţie care codifică pe $G(\mathbb{F})$, dar nu se întâmplă aşa.

Sper că pentru cititor este limpede că apariţia unor erori de acest tip nu împietează asupra adevărului propoziţiei $G(\mathbb{F})$. Mă gândesc, desigur, la *adevărata* $G(\mathbb{F})$, neafectată de erori involuntare sau greşeli de tipar. Îmi amintesc o întâmplare legată de marele fizician american Richard Feynman. Profesorul Feynman îi explica unui student o anumită idee, pe care o formulase defectuos. Când studentul şi-a exprimat nedumerirea, Feynman i-a replicat: „Nu asculta ce-ţi spun, încearcă să pricepi ce am vrut să spun!"[10]

O posibilitate explicită de codificare ar fi să se utilizeze specificaţiile unei maşini Turing, aşa cum am făcut în MN şi, apoi, să se urmărească raţionamentul de tip Gödel, pe care l-am expus în paragraful 2.5 şi pe care l-am codificat în Anexa A. Avem nevoie de o codificare explicită a regulilor din $\mathbb{F}$, în termenii acţiunilor unei anumite maşini Turing, pe care o vom nota cu $T_{\mathbb{F}}$ (proprietatea pe care $T_{\mathbb{F}}$ va trebui să o satisfacă va fi aceia că, dacă unei anumite propoziţii P, constructibilă în termenii limbajului $\mathbb{F}$, i se atribuie numărul natural p, atunci trebuie să ne asigurăm că $T_{\mathbb{F}}(p) = 1$, atunci când P este o teoremă în $\mathbb{F}$, iar $T_{\mathbb{F}}(p)$ nu se opreşte niciodată, pentru celelalte cazuri). Evident, există mari şanse de erori tehnice. În afara dificultăţilor de construire efectivă a lui $T_{\mathbb{F}}$ din $\mathbb{F}$ şi a lui p din P, mai apare şi problema dacă n-am făcut eu o greşeală în specificarea maşinii Turing. Nu cred să fie vreo eroare, dar încrederea mea în mine însumi nu este atât de mare precum încrederea mea în specificaţiile originare ale lui Gödel

[10] Nu pot să localizez cu precizie sursa acestei informaţii. Totuşi, aşa cum a remarcat Richard Jozsa, nu are importanţă dacă m-am exprimat defectuos, căci pot să aplic aici chiar mesajul conţinut în întâmplarea menţionată.

(chiar dacă sunt mai complicate). Sper că a devenit suficient de clar că erorile posibile de acest tip nu sunt esențiale. Trebuie să ținem seama de vorbele lui Feynman!

În ceea ce privește specificațiile mele, există o altă chestiune care ar trebui menționată. În paragraful 2.5 nu mi-am bazat propria versiune a raționamentului Gödel(-Turing) pe consistența lui $\mathbb{F}$, ci pe corectitudinea algoritmului A, ca test al faptului că un anumit calcul nu se termină (adică adevărul propozițiilor-π_1). Acest lucru a mers foarte bine pentru că am văzut că corectitudinea lui A implică adevărul aserțiunii că $C_k(k)$ nu se termină, deci putem utiliza această aserțiune explicită — care este în același timp și o propoziție-π_1 — în locul lui $G(\mathbb{F})$. Mai mult, așa cum am remarcat mai înainte (vezi paragraful 2.8), argumentul depinde de fapt de ω-consistența lui $\mathbb{F}$, nu de consistență. Corectitudinea lui $\mathbb{F}$ implică desigur și ω-consistența lui, exact la fel ca și consistența sa. Nici $\Omega(\mathbb{F})$, nici $G(\mathbb{F})$, nu rezultă din regulile lui $\mathbb{F}$ (conform paragrafului 2.8) presupunând că $\mathbb{F}$ este corect, deși ambele sunt adevărate.

Rezumând, cred că este clar că, oricât ar persista 'umbra de neîncredere' în credințele unui matematician, la trecerea de la credința în corectitudinea unui sistem formal $\mathbb{F}$ la credința în adevărul propozițiilor $G(\mathbb{F})$ (sau $\Omega(\mathbb{F})$), aceasta se datorește în întregime posibilității de a exista erori la formularea (precisă) a lui '$G(\mathbb{F})$' Dar acest lucru nu este relevant pentru discuția de față și n-ar trebui să apară nici o umbră de îndoială asupra versiunii actuale a propoziției $G(\mathbb{F})$. Dacă $\mathbb{F}$ este cu certitudine corect, atunci *acest* $G(\mathbb{F})$ este fără îndoială adevărat. Formele lui $\mathcal{G}$ (sau $\mathcal{G}^{**}$, sau $\mathcal{G}^{***}$) rămân neafectate atunci când 'adevărat' înseamnă 'fără îndoială adevărat'

Q14. Cu siguranță sistemul $\mathbb{ZF}$ — sau o modificare standard a lui $\mathbb{ZF}$ (să-i zicem $\mathbb{ZF}^*$) — reprezintă tot ceea ce este necesar pentru a face matematică serioasă. De ce să nu aderăm la acest sistem, să acceptăm că nu este demonstrabilă consistența sa și să ne vedem de matematica noastră?

Cred că acest punct de vedere este foarte răspândit printre matematicienii activi — în special pentru aceia care nu se preocupă în mod special de fundamentele sau filozofia propriului domeniu de cercetare. Nu este un punct de vedere rezonabil pentru cineva a cărui preocupare principală este să-și vadă de activitatea sa de cercetare matematică (deși, astfel de persoane foarte rar își expun rezultatele în

cadrul unui sistem strict de reguli ca cele ale sistemului $\mathbb{ZF}$). În conformitate cu acest punct de vedere, preocuparea lor este doar în legătură cu ceea ce poate fi demonstrat sau nu în interiorul unui sistem formal ca $\mathbb{ZF}$ (sau o modificare $\mathbb{ZF}^*$). Din acest punct de vedere, obținerea unor rezultate matematice este, de fapt, un fel de *joc*. Să-l botezăm: *jocul* $\mathbb{ZF}$ (sau jocul $\mathbb{ZF}^*$).joc care se cere să fie jucat în conformitate cu acele reguli precise care se află în interiorul sistemului. Acesta este chiar punctul de vedere al formalistului, preocuparea formalistului fiind să afle ce este ADEVĂRAT și ce este FALS și nu în mod necesar ce anume este adevărat și ce anume este fals. Presupunând că sistemul formal este corect, atunci tot ceea ce este ADEVĂRAT va fi și adevărat, iar tot ce este FALS va fi de asemeni fals. Totuși, vor exista unele propoziții, formalizabile în sistem, care să fie adevărate dar nu și ADEVĂRATE, iar unele care, fiind false nu vor fi și FALSE; aceste propoziții, în ambele cazuri, fiind de fapt NEDECIDABILE.

Afirmația lui Gödel[11] $G(\mathbb{ZF})$ și negația sa $\sim G(\mathbb{ZF})$ aparțin acestor două categorii în jocul $\mathbb{ZF}$, presupunând că $\mathbb{ZF}$ este consistent (iar dacă $\mathbb{ZF}$ ar fi inconsistent, atunci atât $G(\mathbb{ZF})$, cât și negația ei $\sim G(\mathbb{ZF})$ ar fi simultan și ADEVĂRATE și FALSE).

Jocul $\mathbb{ZF}$ este, probabil, un punct de plecare perfect îndreptățit pentru a scoate la iveală cele mai multe dintre lucrurile interesante din matematica obișnuită. Totuși, pentru motivul explicat anterior, nu văd cum s-ar putea reprezenta un punct de vedere autentic cu privire la *credințele* matematice ale cuiva. Pentru că, dacă cineva crede că matematica care se face conduce la adevărurile matematice actuale — să zicem la propoziții-π_1 — atunci se poate crede că sistemul folosit este corect; dacă credem că sistemul este corect, atunci se poate crede și că este consistent și astfel trebuie crezută și propoziția-π_1 care afirmă că $G(\mathbb{F})$ este în *realitate* adevărată în ciuda faptului că este NEDECIDABILĂ. Astfel, credințele matematice trebuie să meargă dincolo de ceea ce poate fi dedus în interiorul jocului $\mathbb{ZF}$. Dacă, pe de altă parte, nu există încrederea că $\mathbb{ZF}$ este corect, atunci nu putem crede rezultatele ADEVĂRATE obținute folosind jocul $\mathbb{ZF}$ ca fiind de fapt adevărate.

În orice mod am privi lucrurile, jocul $\mathbb{ZF}$ în sine nu poate reprezenta un punct de pornire cu privire la adevărul matematic (afirmația se aplică în același mod și oricărui $\mathbb{ZF}^*$).

---

[11] Ca mai sus, $\Omega(\mathbb{F})$ se comportă exact la fel ca $G(\mathbb{F})$. Observația se aplică la Q15··Q20.

Q15. Sistemul formal $\mathbb{F}$, pe care alegem să-l folosim, ar putea să nu fie consistent — sau, cel puţin, am putea să nu fim siguri că sistemul $\mathbb{F}$ este consistent, caz în care se pune întrebarea: Pe ce temei putem afirma că $G(\mathbb{F})$ este 'evident' adevărată?

Cu toate că această chestiune a fost amplu tratată în discuţiile precedente, cred că merită să reluăm aici elementele esenţiale, deoarece argumentele de natura lui **Q15** se găsesc în cele mai obişnuite atacuri la adresa modului în care teorema lui Gödel a fost utilizată de Lucas şi de mine.

Trebuie subliniat că noi nu susţinem că $G(\mathbb{F})$ este în mod necesar adevărată pentru orice sistem $\mathbb{F}$, ci că trebuie să tragem concluzia că $G(\mathbb{F})$ este tot atât de adevărată ca orice propoziţie ce se obţine utilizând regulile lui $\mathbb{F}$ (de fapt, $G(\mathbb{F})$ este mai de încredere chiar decât propoziţiile deduse utilizând efectiv regulile din $\mathbb{F}$, deoarece $\mathbb{F}$ ar putea să fie consistent fără ca de fapt să fie şi corect!). Dacă avem încredere în orice propoziţie P, pe care o deducem numai utilizând regulile din $\mathbb{F}$, atunci trebuie să avem încredere şi în $G(\mathbb{F})$, cel puţin până la nivelul la care ne încredem în P. Deci nici un sistem formal $\mathbb{F}$, care poate fi cunoscut — sau algoritmul echivalent F — nu poate reprezenta în totalitate bazele cunoştinţelor noastre adevărate de matematică, sau convingerile noastre. Aşa cum spuneam în comentariile la **Q5** şi **Q6**, argumentaţia este dată sub forma 'reductio ad absurdum': încercăm să presupunem că $\mathbb{F}$ reprezintă în totalitate baza convingerilor noastre şi apoi arătăm că acest lucru duce la o contradicţie, iar de aici rezultă că $\mathbb{F}$ nu poate reprezenta baza pentru convingerile noastre.

Ca în **Q14**, putem să folosim desigur un sistem $\mathbb{F}$ oarecare, chiar dacă nu suntem siguri că el este corect şi deci consistent. Dar dacă există vreun dubiu asupra lui $\mathbb{F}$, atunci, în acest caz, orice rezultat P obţinut utilizând $\mathbb{F}$ trebuie pus sub forma:

'P este deductibil în $\mathbb{F}$'

(adică P este ADEVĂRAT) şi nu să afirmăm pur şi simplu că P este adevărat'. În acest caz avem un enunţ matematic perfect, care poate fi sau adevărat sau fals. Ar fi perfect legitim să ne pronunţăm doar asupra propoziţiilor de acest tip şi, aşa fiind, unii mai fac afirmaţii asupra adevărului matematic absolut. Uneori se poate crede că s-a stabilit că o propoziţie având forma de mai sus este de fapt falsă; adică se poate crede că s-a stabilit

'P nu este deductibil în $\mathbb{F}$'

Propoziţiile de acest tip sunt de forma:

„În acele condiţii... un calcul nu se termină"

(de fapt „ F aplicat lui P nu se termină") şi care are chiar forma propoziţiilor-π_1 considerate anterior. Chestiunea este:

Ce fel de mijloace sunt permise în derivarea propoziţiilor de acest tip?

Care sunt cu adevărat procedurile matematice în care cineva are încredere pentru a stabili adevăruri matematice?

Un astfel de corp de convingeri, dacă sunt rezonabile, nu poate fi echivalent cu încrederea într-un sistem formal, oricare ar fi acela.

Q16. Concluzia că $G(\mathbb{F})$ este adevărată pentru un sistem formal $\mathbb{F}$ depinde de presupunerea că simbolurile din $\mathbb{F}$ folosite pentru a reprezenta numere naturale, chiar reprezintă numere naturale. Pentru alte numere de un tip exotic — să le spunem 'numere supernaturale' — am putea constata că $G(\mathbb{F})$ este falsă. Cum putem şti că în sistemul nostru $\mathbb{F}$ ne referim la numere naturale şi nu la numere supernaturale ?

Este adevărat că nu există nici un mijloc axiomatic finit de a ne asigura că 'numerele' la care ne referim sunt cu adevărat numerele naturale dorite şi nu vreun tip nedorit de numere 'supernaturale' [5]. Dar, într-un anumit sens, acesta este punctul esenţial în discuţia Gödel. Indiferent ce fel de sistem axiomatic $\mathbb{F}$ utilizăm în încercarea de a caracteriza numerele naturale, regulile din $\mathbb{F}$ vor fi insuficiente pentru a ne spune dacă $G(\mathbb{F})$ este, de fapt, adevărată sau falsă. Să presupunem că $\mathbb{F}$ este consistent; ştim că semnificaţia dorită a lui a lui $G(\mathbb{F})$ este ceva cu siguranţă adevărat, nu fals. Totuşi, acest lucru depinde de simbolurile care constituie de fapt expresia formală care este denumită prin $G(\mathbb{F})$ ', având propriile lor semnificaţii. Dacă aceste simboluri sunt reinterpretate ca însemnând altceva decât se presupunea iniţial, putem ajunge la o interpretare ' $G(\mathbb{F})$ ' care să fie cu siguranţă falsă.

Pentru a vedea cum aceste ambiguităţi pot să apară, vom considera noi sisteme formale $\mathbb{F}'$ şi $\mathbb{F}''$, unde $\mathbb{F}'$ se obţine adăugând pe $G(\mathbb{F})$ axiomelor lui $\mathbb{F}$, iar $\mathbb{F}''$ — adăugând axiomelor lui $\mathbb{F}$ pe $\sim G(\mathbb{F})$. Dacă presupunem că $\mathbb{F}$ este corect, atunci $\mathbb{F}'$ şi $\mathbb{F}''$ sunt ambele consistente (căci $G(\mathbb{F})$ este adevărat şi $\sim G(\mathbb{F})$ nu poate fi dedus din regulile lui $\mathbb{F}$).

Dar în interpretarea uzuală a simbolurilor din $\mathbb{F}$ — numită interpretare

standard – presupunând că $\mathbb{F}$ este corect, $\mathbb{F}^*$ va fi şi el corect, dar nu şi $\mathbb{F}^{**}$ Există însă o caracteristică interesantă a sistemelor formale consistente: faptul că se pot găsi reinterpretări ale simbolurilor, în aşa fel încât propoziţiile care erau false în interpretarea standard să fie adevărate în noua interpretare non-standard. Într-o astfel de interpretare non-standard, $\mathbb{F}$ şi $\mathbb{F}^{**}$ pot fi corecte, în loc de $\mathbb{F}^*$ Ne putem imagina că această reinterpretare poate afecta semnificaţia simbolurilor logice (precum '~' şi '&', care în interpretarea standard au sensul de '*non*', respectiv '*şi*'), dar aici avem de a face cu simboluri reprezentând numere nedeterminate ('x', 'y', 'z', etc.) şi semnificaţiile cuantificatorilor logici.

În vreme ce în interpretarea standard '$\forall x$ ' şi '$\exists x$ ' înseamnă 'pentru orice număr natural x respectiv 'există un număr natural x astfel încât într-o interpretare non-standard simbolurile se vor referi la un alt tip de numere, cu proprietăţi de ordine diferite (care ar putea fi numite *numere supernaturale*, folosind terminologia lui Hofstadter (1979)).

Cert este că noi ştim, totuşi, ce sunt numerele naturale şi nu avem nici o dificultate să le distingem de alte feluri ciudate de numere supernaturale. Numerele naturale sunt lucruri obişnuite pe care le notăm în mod normal cu simbolurile 0, 1, 2, 3, 4, 5, 6, Sunt concepte pe care le-am achiziţionat în prima copilărie şi nu avem nici o problemă să le distingem de alte concepte, precum bizarele numere supernaturale (vezi paragraful 1.21). Evident, este ceva misterios în faptul că ştim instinctiv ce sunt de fapt numerele naturale. Copii fiind, sau adulţi, am avut la dispoziţie un număr relativ mic de descrieri legate de ceea ce înseamnă 'zero', 'unu', 'doi', ('trei mere', 'o pară', etc.); şi totuşi prindem întregul concept, în ciuda acestei insuficienţe. Într-un anumit sens platonic, numerele naturale par a fi lucruri cu o existenţă conceptuală absolută, independentă de noi. Cu toată această independenţă, suntem capabili intelectual vorbind – sa intuim şi să intrăm în contact cu conceptul de număr natural, doar cu ajutorul acestor descrieri vagi şi aparent inadecvate. Pe de altă parte, nici un număr finit de axiome nu pot distinge complet numerele naturale de variantele posibile pe care le-am numit 'numere supernaturale'.

Mai mult, noi suntem în stare să percepem direct caracterul infinit al totalităţii numerelor naturale, în vreme ce un sistem constrâns să opereze cu un număr finit de reguli precise nu poate distinge caracterul infinit particular al numerelor naturale de al altor posibile numere supernaturale. Infinitatea care

caracterizează numerele naturale o înţelegem chiar atunci când o reprezentăm numai prin puncte, puncte, ' în descrierea 0, 1, 2, 3, 4, 5, 6, ', sau prin 'etc.' în 'zero, unu, doi, trei, etc.'

Nu am spus exact ce înseamnă — ce este un număr natural în termenii unor reguli precise. Din fericire, pentru că nici nu este posibil. Într-un fel sau altul, am descoperit că *ştim* ce este un număr natural de îndată ce ne-am îndreptat în direcţia bună!

Unii dintre cititori cunosc poate axiomele lui Peano pentru aritmetica numerelor naturale (la care am făcut aluzie în treacăt în paragraful 2.7) şi s-ar putea să fie nedumeriţi de ce ele nu definesc suficient de bine numerele naturale. În concordanţă cu definiţia lui Peano, pornim cu un simbol 0 şi cu un operator de succesiune notat cu S, interpretat ca simpla adunare cu 1 (la numărul cu care se operează, astfel că definim pe 1 prin operaţia S0, iar 2 va fi S1 adică SS0, etc. Avem ca reguli, faptul că dacă $Sa = Sb$ atunci $a = b$; şi că, pentru nici un x, Sx nu poate fi 0 Deasemeni, există 'principiul inducţiei' care susţine că o proprietate P a numerelor trebuie să fie adevărată pentru toate numerele n , dacă satisface următoarele două condiţii:

(i) dacă $P(n)$ este adevărată, atunci şi $P(Sn)$ este de asemeni adevărată, pentru orice n ;

(ii) $P(0)$ este adevărată.

Neplăcerea începe cu operaţiile logice, unde, în interpretarea standard, simbolurile $\forall$ şi $\exists$ numesc, aşa cum am mai văzut, expresiile 'oricare ar fi *numărul natural*' şi 'există *numărul natural* astfel încât' Într-o interpretare non-standard, semnificaţiile acestor simboluri se va modifica corespunzător, astfel încât ele vor cuantifica alte tipuri de numere, în locul celor naturale. În vreme ce specificările matematice Peano, care sunt date pentru operaţia de succesiune S, caracterizează efectiv relaţia de ordine care distinge numerele naturale de alte tipuri de numere, aceste specificări nu sunt capturate în termenii regulilor formale pe care aceşti cuantificatori le satisfac. Cu scopul de a captura semnificaţiile specificaţiilor lui Peano, trebuie să trecem la ceea ce este cunoscut ca 'logica de ordinul doi' în care sunt introduşi cuantificatori precum $\forall$ şi $\exists$, dar în care domeniul lor de aplicare îl formează mulţimi infinite de numere naturale şi nu doar un singur număr natural.

În 'logica de ordinul întâi' a aritmeticii lui Peano cuantificatorii acţionează pe câte un număr şi obţinem un sistem formal în sens obişnuit. Dar logica de

ordinul doi nu ne furnizează un sistem formal. Într-un sistem formal strict trebuie să existe o metodă pur formală (adică algoritmică) pentru a decide dacă regulile sistemului formal au fost corect aplicate, ceea ce, în orice caz, constituie un motiv esențial pentru folosirea sistemelor formale în contextul de față. Această proprietate nu mai este prezentă în logica de ordinul doi. Este o concepție greșită, dar foarte comună, în spiritul sentimentelor exprimate în **Q16**, că teorema lui Gödel arată că există multe feluri de aritmetici, fiecare în parte la fel de valabilă. Aritmetica particulară, pe care o alegem să lucrăm, va fi definită, în consecință, pur și simplu de un sistem formal arbitrar ales. Teorema lui Gödel arată că nici unul din aceste sisteme formale, dacă este consistent, nu poate fi complet; se argumentează că astfel am putea continua să adăugăm noi axiome, după capriciile noastre și să obținem tot felul de sisteme alternative consistente, în interiorul cărora să dorim să lucrăm. Se face uneori comparația cu situația la care s-a ajuns în geometria euclidiană. Timp de 21 de secole s-a crezut că geometria euclidiană era singura geometrie posibilă. Dar când, în secolul nouăsprezece, matematicieni precum Gauss, Lobacevsky și Bolyai, au arătat că există de fapt alternative egal posibile — problematica din geometrie a părut că se mută din absolut în arbitrar. Tot astfel, adeseori se susține că Gödel a arătat că și aritmetica este o chestiune de alegere arbitrară, oricare mulțime de axiome consistente fiind la fel de bună ca alta.

Acest mod de prezentare este o interpretare complet greșită a rezultatelor pe care Gödel le-a demonstrat. El ne-a învățat că noțiunea de sistem formal este, în fond nepotrivită pentru a captura conceptele matematice, chiar cele mai simple. Atunci când folosim termenul de 'aritmetică' fără altă calificare, înțelegem chiar aritmetica obișnuită care operează cu numerele naturale obișnuite și nu cu alt fel de numere 'supernaturale'. Dacă dorim, putem să studiem proprietățile sistemelor formale și acest lucru este, cu siguranță, o parte valoroasă a strădaniei din matematică. Dar acest lucru este oarecum diferit de explorarea proprietăților obișnuite ale numerelor naturale. Situația este, într-un anumit fel, nu chiar atât de diferită de cea care a intervenit în geometrie. Studiul geometriilor neeuclidiene este interesant din punct de vedere matematic, cu aplicații importante (ca cele din fizică, vezi MN, cap. 5 — în special figurile 5.1, 5.2, dar și paragraful 4.4), dar, atunci când termenul 'geometrie' este folosit în limbajul obișnuit (diferit de situația în care un matematician sau un specialist în fizică teoretică l-ar putea folosi) înțelegem geometria obișnuită a lui Euclid. Există totuși o diferență și anume: ceea ce un

logician ar înţelege prin 'geometrie euclidiană' poate fi specificat (cu anumite rezerve [6]) în termenii unui sistem formal particular, în vreme ce, aşa cum a arătat Gödel, 'aritmetica' obişnuită nu poate fi specificată.

Mai degrabă decât să arate că matematica (mai exact aritmetica) este o preocupare arbitrară, a cărei direcţie este hotărâtă de capriciul Omului, Gödel a demonstrat că este ceva absolut, ce trebuie descoperit şi nu inventat (vezi paragraful 1.17). Am văzut chiar noi ce sunt numerele naturale şi că nu avem probleme să le distingem de oricare fel de numere supernaturale. Gödel a arătat că nici un sistem de reguli 'făcute de om' nu poate el însuşi să obţină acest lucru pentru noi. Un astfel de punct de vedere platonic a fost important pentru Gödel şi este important şi în viitoarele consideraţii din această carte (paragraful 8.7).

Q17. Să presupunem că sistemul formal $\mathbb{F}$ este proiectat să reprezinte adevărurile matematice, în principiu accesibile minţii. Oare nu am putea ocoli problema incapacităţii de a încorpora propoziţia lui Gödel $G(\mathbb{F})$ în mod formal în $\mathbb{F}$, prin încorporarea a ceva cu *semnificaţia* lui $G(\mathbb{F})$, folosind o reinterpretare a sensurilor simbolurilor din $\mathbb{F}$?

Există într-adevăr căi de reprezentare a raţionamentului lui Gödel aplicat lui $\mathbb{F}$ în sistemul formal $\mathbb{F}$ (suficient de extins), aşa încât semnificaţiile simbolurilor din $\mathbb{F}$ sunt reinterpretate ca ceva diferit de semnificaţiile atribuite iniţial simbolurilor din $\mathbb{F}$. Cu toate acestea, pare o adevărată înşelătorie să încercăm să interpretăm $\mathbb{F}$ ca procedura prin care mintea ajunge la concluziile sale matematice. Simbolurile din $\mathbb{F}$ trebuie să nu poată să-şi schimbe semnificaţiile dacă activităţile mentale sunt interpretate numai în termenii lui $\mathbb{F}$.

Dacă activitatea mentală poate conţine ceva mai presus de operaţiile din $\mathbb{F}$ propriu zise, de exemplu, schimbarea semnificaţiei acelor simboluri, atunci trebuie să cunoaştem regulile care guvernează detaliile schimbării acestor semnificaţii. Sau aceste reguli sunt nealgoritmice, caz în care $\mathcal{G}$ se confirmă, sau există un procedeu algoritmic, caz în care va trebui să încorporăm *acest* procedeu pe primul loc în '$\mathbb{F}$ — să numim acest sistem $\mathbb{F}^+$ — pentru ca el să reprezinte totalitatea intuiţiilor noastre şi atunci n-ar fi fost necesar ca semnificaţiile să se schimbe.

În acest ultim caz propoziţia $G(\mathbb{F}^+)$ ia locul propoziţiei $G(\mathbb{F})$ din discuţia precedentă şi astfel n-am câştigat de fapt nimic.

Q18. **Se poate formula, într-un sistem atât de simplu ca aritmetica lui Peano, o teoremă a cărei interpretare să aibă implicaţia:**

' $\mathbb{F}$ corect ' implică '$G(\mathbb{F})$'

Nu este acest lucru tot ceea ce doream de la teorema lui Gödel? Cu siguranţă ne-ar permite să trecem de la încrederea în corectitudinea oricărui sistem formal $\mathbb{F}$ la încrederea în adevărul propoziţiei Gödel asociate, atâta timp cât suntem pregătiţi să acceptăm chiar aritmetica lui Peano?

Este adevărat desigur că o astfel de teoremă [7] poate fi formulată în aritmetica lui Peano. Mai precis (cum nu putem încapsula în mod corespunzător noţiunea de 'corectitudine' şi 'adevăr' în interiorul nici unui sistem formal — aşa cum rezultă dintr-o teoremă celebră a lui Tarski), ceea ce se poate de fapt formula este un rezultat mai tare:

' $\mathbb{F}$ consistent ' implică '$G(\mathbb{F})$ '

sau, altfel

' $\mathbb{F}$ ω-consistent ' implică '$\Omega(\mathbb{F})$ '

Acestea au implicaţia cerută pentru **Q18** pentru că dacă $\mathbb{F}$ este corect el este cu siguranţă consistent sau ω-consistent, după cum e cazul. Cu condiţia să înţelegem *semnificaţia* simbolurilor utilizate, putem într-adevăr să trecem de la încrederea în corectitudinea lui $\mathbb{F}$ la încrederea în adevărul lui $G(\mathbb{F})$. Dar acest lucru a fost deja acceptat. Dacă înţelegem semnificaţiile atunci putem trece cu adevărat de la $\mathbb{F}$ la $G(\mathbb{F})$ în mod *automat*. Dacă asta ar fi posibil am putea automatiza procedura generală de gödelizare şi am putea construi un dispozitiv automat care într-adevăr ar conţine tot ceea ce cerem teoremei lui Gödel. Dar acest lucru nu poate fi făcut; căci dacă am adăuga această presupusă procedură algoritmică unui sistem formal oarecare $\mathbb{F}$, am obţine, ca efect doar un nou sistem formal $\mathbb{F}'$ şi propoziţia sa asociată $G(\mathbb{F}')$ s-ar afla în afara lui $\mathbb{F}'$ Rămân astfel întotdeauna *unele* aspecte ale cunoaşterii oferite de teorema lui Gödel, oricât de mult s-ar fi încorporat din ea în procedura formalizată sau algoritmizată. Această 'înţelegere gödeliană' cere o continuă referire la semnificaţia actuală a simbolurilor oricărui sistem căreia i se aplică procedura Gödel. În acest fel, dificultatea cu **Q18** este asemănătoare cu cea ridicată de **Q17**. Faptul că gödelizarea nu poate fi automatizată este strâns legată de discuţia de la **Q6** şi de cea de la **Q19** care urmează.

Există un alt aspect al lui **Q18** care merită luat în considerare. Să ne

imaginăm că avem un sistem formal corect $\mathbb{E}$ care include aritmetica lui Peano. Teorema la care ne-am referit în **Q18** va fi printre implicaţiile lui $\mathbb{E}$ şi o instanţiere particulară a sa, care se aplică unui $\mathbb{F}$ particular care este chiar $\mathbb{E}$, va fi o teoremă a lui $\mathbb{E}$ Astfel putem spune că una dintre implicaţiile lui $\mathbb{E}$ este:

' $\mathbb{E}$ corect' implică '$G(\mathbb{E})$ ';

sau mai precis:

' $\mathbb{E}$ consistent' implică '$G(\mathbb{E})$ ';

Dar aceste aserţiuni ar trebui, în termenii semnificaţiilor lor actuale să aducă după ele implicaţia că $G(\mathbb{E})$ este, de fapt, atestată. Deoarece — cu privire la prima dintre aserţiunile de mai sus — orice afirmaţie pe care o face $\mathbb{E}$ este condiţionată de presupunerea că $\mathbb{E}$ este corect; astfel, dacă $\mathbb{E}$ afirmă ceva care este în mod explicit condiţionat de propria corectitudine, poate de asemenea să afirme direct acest lucru (afirmaţia 'dacă eu sunt crezut atunci X este adevărat' îndreptăţeşte afirmaţia simplă făcută de acelaşi vorbitor, 'X este adevărat!'). Totuşi, un sistem formal corect $\mathbb{E}$ *nu poate* afirma $G(\mathbb{E})$ ceea ce reflectă faptul că este incapabil să afirme propria sa corectitudine. Mai mult, vedem că el nu poate, de fapt, să încapsuleze semnificaţia simbolurilor cu care operează. Acelaşi lucru se vede şi în legătură cu cea de-a doua dintre aserţiunile de mai înainte, dar cu adaosul ironic că în vreme ce $\mathbb{E}$ este incapabil să-şi afirme propria consistenţă când de fapt el este consistent, nu suferă de o astfel de inhibiţie când este *inconsistent*. Un sistem inconsistent $\mathbb{E}$ poate afirma drept teoremă orice lucru pe care l-ar putea formula! Asta ne conduce la formularea ' $\mathbb{E}$ este consistent' Un sistem (suficient de extins) va atesta propria consistenţă dacă şi numai dacă este *inconsistent*!

Q19. De ce nu adoptăm pur şi simplu o procedură de alăturare repetată a propoziţiei Gödel $G(\mathbb{F})$ la orice sistem formal $\mathbb{F}$ pe care îl acceptăm şi să permitem acestei proceduri să continue la infinit?

Fiind dat un sistem formal particular $\mathbb{F}$, care este suficient de extins şi presupus corect, putem imagina un mod în care să asociem $G(\mathbb{F})$ lui $\mathbb{F}$ ca o nouă axiomă şi astfel să obţinem un nou sistem $\mathbb{F}_1$, care este şi el perceput ca fiind corect. Pentru a avea o notaţie unitară, în cele ce urmează, putem scrie $\mathbb{F}_0$ pentru $\mathbb{F}$). Acum, vom asocia pe $G(\mathbb{F}_1)$ lui $\mathbb{F}_1$ şi astfel obţinem un nou sistem $\mathbb{F}_2$, la fel perceput ca corect. Acum, cu un efort suplimentar, vom fi capabili să vedem cum să construim încă un sistem formal $\mathbb{F}_\omega$ a cărui axiome să includă întreaga

mulţime infinită { $G(\mathbb{F}_0)$, $G(\mathbb{F}_1)$, } ca axiome adiţionale la $\mathbb{F}$. Acest sistem $\mathbb{F}_\omega$ va fi de asemeni corect. Putem continua procesul adăugând pe $G(\mathbb{F}_\omega)$ la $\mathbb{F}_\omega$ pentru a-l obţine pe $\mathbb{F}_{\omega+1}$, apoi pe $G(\mathbb{F}_{\omega+1})$ la $\mathbb{F}_{\omega+1}$ pentru a-l obţine pe $\mathbb{F}_{\omega+2}$ şi aşa mai departe. Apoi, la fel ca mai înainte, putem încorpora întreaga mulţime infinită de axiome şi să obţinem astfel pe $\mathbb{F}_{\omega+\omega}$ care este şi el corect.

Procesul poate fi continuat, de la indicele $\omega_2 = \omega+\omega$ trecând, prin ω_2+1, la $\omega_3 = \omega_2+\omega$ şi tot aşa, pentru a forma un sistem $\mathbb{F}_{\omega\omega}$, etc.[12]

Cititorii familiarizaţi cu notaţia lui Cantor pentru numerele ordinale au recunoscut deja indexarea pe care am folosit-o ca notaţie pentru astfel de numere ordinale. Cei ce nu sunt familiarizaţi cu astfel de lucruri nu trebuie să se neliniştească. Este suficient să spunem că această procedură de gödelizare poate fi continuată la infinit, către ordinalé tot mai mari, ca ω^ω ..., la fiecare pas putând vedea cum să sistematizăm întreaga mulţime de gödelizări pe care le-am obţinut până în acel punct. Aceasta este de fapt dificultatea problemei: ceea ce am numit anterior prin cuvintele 'un efort suplimentar' presupunând de fapt înţelegerea potrivită a modului în care se sistematizează gödelizarea. Sistematizarea este posibil să fie atinsă cu condiţia ca ordinalul care a fost atins să fie etichetat ca ordinal *recursiv*, ceea ce înseamnă, de fapt, că există un algoritm de generare a procedurii. Totuşi, nu există nici o procedură algoritmică care să ne permită să facem această sistematizare pentru *toate* ordinalele recursive, odată pentru totdeauna. Trebuie să continuăm să folosim din nou intuiţia.

Procedeul tocmai descris a fost pentru prima oară formulat de Alan Turing în teza sa de doctorat (publicată în Turing 1939) [8] şi arată că există un sens în care *orice* propoziţie-π_1 adevărată poate fi demonstrată prin gödelizare repetată de tipul descris aici (vezi Feferman 1988). Totuşi acest lucru nu ne oferă o procedură matematică pentru a stabili adevărul propoziţiilor-π_1 chiar pentru motivul că nu se poate sistematiza mecanic gödelizarea. De fapt, noi putem *deduce*, chiar din rezultatul lui Turing, că gödelizarea nu poate fi făcută mecanic. Deoarece am stabilit deja (efectiv, în paragraful 2.5) că adevărul propoziţiilor-π_1 nu poate fi decis de nici o procedură algoritmică. Dar repetarea gödelizării nu ne permite să obţinem nimic pe calea unei proceduri de sistematizare care se află în afara

[12] Autorul continuă 'gödelizarea' aceasta, repetând procedeul pentru $\omega_3 = \omega_2+\omega$, ω_4, ω_5,... , pentru a trece apoi lá $\omega^2 = \omega\omega$, ω^3, ω^4, ... pentru a ajunge la ω^ω, apoi la $(\omega^\omega)^\omega$, . Ne-am permis să scurtăm puţin acest pasaj (n.t.)

considerațiilor de calculabilitate de care ne-am ocupat până acum. Astfel, **Q19** nu reprezintă nici o amenințare la adresa lui $\mathcal{G}$

Q20. Să fie, oare, valoarea reală a înțelegerii matematice faptul că ne permite să ajungem la lucruri necalculabile sau că ne permite să înlocuim calcule foarte complicate prin intuiții relativ simple? Cu alte cuvinte, nu suntem oare în situația în care mintea ne permite să găsim căi mai scurte în raport cu teoria complexității, mai degrabă decât să trecem dincolo de marginile a ceea ce este calculabil ?

Sunt gata să cred că în *practică* intuițiile unui matematician sunt mult mai adesea implicate în 'surprinderea' complexității calculului decât în chestiunea necalculabilității. La urma urmelor, matematicienii au tendința să fie niște oameni leneși, încercând adesea să găsească mijloace de a evita calculele (chiar dacă asta îi duce la un efort mental mult mai mare față de calculul propriu-zis!). Adeseori se întâmplă că încercarea de a face calculatoarele să 'macine' în mod mecanic pentru a scoate niște teoreme, într-un sistem formal destul de simplu, conduce la o fundătură de calcule complicate — în vreme ce, înarmat cu înțelegerea adecvată a semnificației regulilor sistemului, un om care se ocupă de matematică nu va avea prea mult de furcă cu deducerea unor rezultate interesante din acel sistem [9].

M-am ocupat de necalculabilitate, în argumentația mea, mai degrabă decât de complexitate, pur și simplu pentru motivul că numai cu necalculabilitatea am fost în stare să văd cum formulez afirmațiile tari necesare. Se poate foarte bine întâmpla însă ca, în activitatea celor mai mulți matematicieni, chestiunea necalculabilității să joace, dacă nu nici un rol, în orice caz, unul foarte mic. Dar nu aceasta este chestiunea în discuție. Voi încerca să arăt că înțelegerea matematică se află dincolo de calcul, iar raționamentul Gödel (Turing) este una dintre puținele șanse pe care le avem. Este foarte probabil ca intuițiile noastre matematice și înțelegerea să fie utilizate adeseori pentru a obține ceva ce s-ar putea obține în principiu și prin calcul — dar calculul orb, fără multă pătrundere, se va dovedi atât de ineficient, încât să nu fie utilizabil (conform paragrafului 3.26). Însă, aceste aspecte sunt mult mai dificil de tratat decât chestiunea necalculabilității.

În orice caz, oricât de adevărate ar fi sentimentele exprimate de **Q20**, acestea nu contrazic în nici un fel pe $\mathcal{G}$.

Note şi referinţe bibliografice

1. Ar putea să pară că acesta este perfect 'evident' şi nu un subiect de dispută între matematicieni. Totuşi, problema apare în legătură cu noţiunea de 'existenţă' a mulţimilor infinite (vezi, spre exemplu, Smorynski 1975, Rucker 1984, Moore 1990). Am văzut din exemplul paradoxului lui Russell că trebuie să fim foarte atenţi cu acest tip de probleme. În conformitate cu un anumit punct de vedere, o mulţime nu ar trebui considerată că există decât dacă există cel puţin câteva reguli clare (nu neapărat calculabile), pentru specificarea elementelor care aparţin mulţimii şi a celor care nu aparţin. Este chiar ceea ce axioma alegerii nu ne oferă, căci nu există — nu se dă — vreo regulă pentru a specifica care element trebuie luat din fiecare membru al colecţiei (unele din implicaţiile axiomei alegerii sunt foarte neintuitive şi aproape paradoxale. Probabil că acesta este unul din motivele pentru care axioma alegerii este atât de controversată. Nici eu nu sunt chiar sigur asupra poziţiei mele în această chestiune!)).
2. În capitolul final al cărţii sale din 1966, Cohen face observaţia că, deşi a arătat că ipoteza continuului este NEDECIDABILĂ, în conformitate cu procedurile lui ZF, el a lăsat neatinsă întrebarea dacă ea este de fapt adevărată sau nu şi discută cum s-ar putea ajunge să se decidă asupra acestei întrebări. Prin urmare este clar că el nu împărtăşeşte părerea cum că ar fi complet arbitrar dacă ipoteza continuului este sau nu acceptată. Asta în contradicţie cu opiniile deseori exprimate despre implicaţiile rezultatelor Gödel-Cohen şi anume că există numeroase 'teorii alternative a mulţimilor' care sunt la fel de 'valabile' pentru matematică. Prin aceste observaţii, Cohen se prezintă pe sine ca fiind, la fel cu Gödel, un adevărat platonician pentru care problemele adevărului matematic sunt *absolute* şi nu arbitrare. Acest lucru corespunde pe deplin cu propriile mele vederi.

   Vezi, de exemplu, Hofstadter (1981), Bowie (1982).
4. De exemplu, vezi comentariile din *Behavioral and Brain Sciences*, **13** (1990), 643-705.
5. Această terminologie a fost sugerată de Hofstadter (1981). Este o 'altă' teoremă a lui Gödel — teorema de completitudine — care ne spune că astfel de modele non-standard există totdeauna.
6. De fapt, depinde care propoziţii sunt considerate ca parte a ceea ce se numeşte aici 'Geometrie euclidiană'. În terminologia uzuală a logicienilor, sistemul 'geometriei lui Euclid' ar trebui să includă propoziţii de un tip special şi rezultă că adevărul sau falsitatea acestor propoziţii ar trebui rezolvat în termenii unei proceduri algoritmice — deci afirmaţia că geometria lui Euclid poate fi specificată în termenii unui sistem formal. Totuşi, în alte interpretări, 'aritmetica' obişnuită

poate fi şi ea considerată ca o parte a 'geometriei euclidiene' şi aceasta admite clase de propoziţii care nu pot fi soluţionate algoritmic. Acelaşi lucru se aplică dacă considerăm că problema de acoperire este o parte a geometriei euclidiene — ceea ce pare foarte natural. În acest sens, geometria euclidiană nu poate fi specificată formal 'cu nimic mai bine' decât aritmetica!

7. Vezi comentariul lui Davis (1991).
8. Vezi de asemeni Kreisel (1960, 1967), Good (1967).
9. Vezi Freedman (1994) relativ la unele probleme ce pot apare în sistemele de calcul la încercarea de realizare a unei matematici 'proprii' acestora. În general, astfel de sisteme nu ajung prea departe. Au o nevoie stringentă de un ghid uman!

Anexa A

O maşină Turing de gödelizare

Presupunem dată o procedură algoritmică A care determină cu precizie că anumite calcule nu se termină. Voi oferi un procedeu explicit de construire, din A, a unui calcul particular C pentru care A greşeşte; adică vom vedea că C nu se termină. Luând această expresie explicită a lui C, putem examina gradul său de complicaţie, comparându-l cu cel al lui A, aşa cum este necesar pentru argumentele din paragrafele 2.6 (conform cu Q8) şi 3.20.

Pentru definiţie, voi apela la specificarea unei maşini Turing particulare, aşa cum este dată în **MN** în capitolul 2. Pentru detalii complete cititorul este îndrumat spre volumul menţionat. Aici voi da doar o descriere minimală adecvată scopului.

O maşină Turing are un număr finit de stări interne dar acţionează asupra unei benzi infinite. Banda constă dintr-o succesiune de spaţii (cadre), unde fiecare cadru poate fi sau marcat sau nemarcat şi pot apare un număr total finit de marcaje pe bandă. Să notăm fiecare cadru marcat cu 𝟙 iar fiecare cadru nemarcat cu 𝟘. Există un dispozitiv de citire care examinează câte un marcaj în unitatea de timp şi, funcţie de starea internă a maşinii Turing şi de marcajul tocmai citit, se determină trei elemente: (i) dacă marcajul tocmai citit va fi alterat sau lăsat aşa cum este, (ii) care va fi noua starea internă a maşinii, (iii) dacă dispozitivul de citire va citi la pasul următor o poziţie din dreapta (notat cu **R**) sau din stânga (notat cu **L**) (deplasarea benzii), sau o poziţie la dreapta împreună cu oprirea (stare notată cu STOP). Când maşina ajunge să se oprească, atunci pe ecranul ei apare afişat rezultatul calculului care s-a efectuat, ca succesiune de cifre binare (𝟘 şi 𝟙), în stânga dispozitivului de citire. Iniţial, banda este complet albă, exceptând

marcajele ce constituie datele inițiale (ca set finit de cifre binare). La început, dispozitivul de citire este situat în stânga tuturor marcajelor de pe bandă.

Când reprezentăm numere naturale pe bandă, fie ca date inițiale fie ca rezultate, este util să se folosească notația binară extinsă, în concordanță cu care numerele se scriu efectiv ca și coduri binare, dar cifra '1' este înlocuită cu 10, în timp ce '0' este scrisă 0 Astfel, avem următoarea schemă de transformare a numerelor în notația binar extinsă:

0	0	10	100100	20	1001000
1	10	11	1001010	21	10010010
2	100	12	101000	22	10010100
3	1010	13	1010010	23	100101010
4	1000	14	1010100	24	10100000
5	10010	15	10101010	25	10100010
6	10100	16	100000	26	10100100
7	101010	17	1000010	27	101001010
8	10000	18	1000100	28	10101000
9	100010	19	10001010	29	101010010

Subliniem că în notația binară expandată nu poate apărea niciodată 1 în două poziții succesive. În consecință, putem semnala începutul și sfârșitul specificării unui număr natural printr-o succesiune de doi sau mai mulți de 1 Atunci putem folosi secvențele 110, 1110, 11110, etc., de pe bandă pentru a nota (codifica) diferite tipuri de instrucțiuni.

Putem de asemeni să folosim marcajele de pe bandă pentru a specifica mașini Turing particulare. Aceasta este necesar când considerăm acțiunea unei mașini Turing universale U.

Mașina universală U acționează asupra unei benzi a cărei porțiune inițială oferă specificațiile detaliate ale unei mașini Turing particulare T pe care mașina universală tinde s-o imite. Datele asupra cărora T ar acționa sunt transmise mașinii U , pe bandă, la dreapta porțiunii care determină mașina T. Pentru a specifica

maşina T, putem folosi secvenţele 110, 1110, şi 11110 pentru a nota, respectiv, diferitele instrucţiuni pentru dispozitivul de citire a lui T, mai precis deplasarea benzii, cu un pas la dreapta, un pas la stânga, sau oprirea după deplasarea cu un pas la dreapta:

R ↔ 110
L ↔ 1110
STOP ↔ 11110

Simbolul 0 sau secvenţa 10 va precede fiecare astfel de instrucţiune pentru a indica că că dispozitivul de citire va marca banda cu respectiv, 0 sau 1, în locul unde se află simbolul tocmai citit. Imediat înaintea simbolurilor menţionate mai sus, 0 sau 10, va fi expresia binară expandată a numărului reprezentând starea internă în care maşina Turing trebuie trecută, în concordanţă cu această instrucţiune (stările interne, fiind finite ca număr, pot fi etichetate cu numerele naturale consecutive 0, 1, 2, 3, 4, 5, 6,..., N; în codurile de pe bandă, notaţia binar expandată va fi folosită pentru aceste numere).

Instrucţiunea particulară, la care această operaţie se referă, va fi determinată prin starea internă a maşinii chiar înainte de citirea de pe bandă, împreună cu simbolul 0 sau 1, pe care dispozitivul de citire tocmai urmează să-l citească şi probabil să-l schimbe. Spre exemplu, ca parte a specificării lui T, pot apare instrucţiuni precum 230 → 171R, care înseamnă 'dacă T este în starea internă 23 şi dispozitivul de citire întâlneşte un 0 pe bandă, atunci înlocuieşte acest semn prin 1, trece în starea internă 17 şi deplasează banda cu un pas spre dreapta' În acest caz, partea '171R' din instrucţiune se va codifica prin 10000101011 0. Segmentând acest şir în forma 100001 0.10.110, vom vedea că prima porţiune este numărul 17 scris în binar expandat, a doua porţiune codifică marcarea lui 1 pe bandă, iar a treia parte codifică instrucţiunea 'deplasare la dreapta' Cum putem specifica starea internă precedentă (aici starea internă 23) şi marcajul examinat de pe bandă (aici 0)? Dacă se doreşte, putem explicita acestea în termenii respectivelor numere binar expandate. Desigur, asta nu este necesar de fapt, căci ordonarea diferitelor instrucţiuni ar fi suficientă (adică ordonarea 00 →, 01 →, 10 →, 11 →, 20 →, 21 →, 30 →, ...).

Acestea ne furnizează în esenţă codurile maşinilor Turing date în MN, dar pentru completare, câteva puncte adiţionale merită menţionate. În prima poziţie,

ne vom asigura că există o instrucţiune dată pentru fiecare din stările interne, acţionând prin 0 şi 1 (exceptând faptul că o instrucţiune pentru starea internă cu numărul cel mai mare acţionând asupra lui 1, nu este deloc necesară). Când nu există nici o variantă de folosire a unei astfel de instrucţiuni în program, atunci se inserează o instrucţiune neexecutabilă. Spre exemplu, o astfel de instrucţiune poate fi 231 → 00R, dacă se întâmplă ca starea internă 23 să nu întâlnească niciodată marcajul 1 în execuţia programului.

În specificarea codificată (pe o bandă) a unei maşini Turing, în concordanţă cu indicaţiile de mai sus, perechea 00 poate fi reprezentată prin secvenţa 000, dar se poate, din economie, utiliza un singur 0, fără ambiguitate, pentru a separa secvenţele de unu sau mai mulţi 1, de ambele părţi[1]

Maşina Turing începe să lucreze fiind în starea internă 0, iar dispozitivul de citire parcurge banda, ramânând în aceeaşi stare, până când întâlneşte primul 1. Aceasta se obţine prin presupunerea că operaţia 00 → 00R este întotdeuna parte a instrucţiunilor unei maşini Turing. Astfel, în specificarea actuală a unei maşini Turing ca o succesiune de 0-uri şi 1-uri, nu mai este necesar să prevedem explicit această instrucţiune; în loc de asta, pornind cu 01 → X, unde X desemnează prima operaţie netrivială a maşinii puse în funcţiune (adică prima întâlnire a unui 1 de pe bandă). Aceasta sugerează ştergerea secvenţei iniţiale 110 (notând → 00R) care altfel va apare mereu în secvenţele specificând maşina Turing. Mai mult, vom şterge de asemeni şi secvenţa finală 110 din această specificare, pentru că şi aceasta este comună tuturor maşinilor Turing.

Secvenţa care rezultă, de cifre 0 şi 1 constituie chiar codificarea binară (ordinară, ne-expandată) a numărului n al maşinii Turing (pentru maşina în discuţie, aşa cum a fost dat şi în MN, capitolul 2). Vom vorbi deci despre a n-a maşină Turing, $T = T_n$ Oricare astfel de număr binar, când se adaugă o secvenţă

---

[1] Asta înseamnă că în codificarea unei maşini Turing, fiecare apariţie a secvenţei ...110011... poate fi înlocuită prin ...11011... Există 15 locuri în specificarea maşinii Turing universale, specificare dată de mine în MN (vezi nota 7 de la cap.2), unde am omis să fac această înlocuire. Acesta este un serios motiv de iritare în ce mă priveşte, deoarece am trudit considerabil cu cerinţele de specificare, pe care le impusesem, pentru a găsi un număr pentru această maşină Turing universală, număr suficient de mic ca să poată fi dedus. Substituţiile menţionate mai sus ar fi furnizat un număr de aproape 30000 de ori mai mic decât cel găsit de mine! Ţin să-i mulţumesc lui Steven Gunhouse pentru observaţia aceasta pe care mi-a comunicat-o, cât şi pentru verificare întreprinsă independent, că specificarea, aşa cum apare tipărită în MN, defineşte efectiv o maşină Turing universală.

110 la sfârşit, este o înşiruire de 0-uri şi 1-uri, în care nu apar niciodată mai mult de 4 de 1 unul după altul. Un număr n care nu respectă această condiţie ne va da o 'falsă maşină Turing' care îşi va suspenda execuţia ori de câte ori va întâlni 'instrucţiuni' cuprinzând mai mult decât 4 cifre 1. O astfel de maşină T_n vom spune că este *incorect specificată*. Acţiunea ei asupra oricărei benzi va fi, prin *definiţie*, fără sfârşit.

Similar, dacă o maşină Turing în acţiune întâlneşte o instrucţiune care ar introduce o stare specificată de un număr mai mare decât toate cele prin care au fost listate celelalte instrucţiuni, atunci avem tot o eroare; şi vom considera că avem tot o maşină falsă şi că acţiunea ei nu se termină (s-ar putea elimina acest inconvenient fără prea mare dificultate prin utilizarea altor instrumente, dar nu este de fapt nevoie de asta, conform paragrafului 2.6, **Q4**).

Pentru a vedea cum să construim, dintr-un algoritm dat A, calculul pentru care A să eşueze, trebuie să presupunem că A este dat ca o maşină Turing. Această maşină acţionează asupra benzii codificând două numere naturale p şi q. Trebuie să admitem că dacă $A(p,q)$ este un calcul care se termină, atunci acţiunea calculului T_p asupra numărului q nu se încheie niciodată. Amintim că dacă T_p nu este corect specificată, atunci trebuie să considerăm că acţiunea sa asupra lui q nu se termină, oricare ar fi q. Pentru orice astfel de p 'prohibit', orice rezultat pentru $A(p,q)$ va fi consistent cu presupunerile noastre. În consecinţă, trebuie să avem de-a face doar cu numere p pentru care T_p este corect specificată. Atunci, expresia binară a numărului p, aşa cum se reprezintă pe bandă, nu poate conţine secvenţe ...11111... Exact asta ne permite să folosim secvenţa 11111 pentru a marca începutul şi sfârşitul numărului p, aşa cum apare el pe bandă.

Totuşi, ar trebui să facem acelaşi lucru şi cu q, care însă nu-i un număr neapărat de acelaşi tip cu p. Asta ne aduce câteva dificultăţi tehnice, pentru prescrierile maşinii Turing şi va fi mai convenabil să le ocolim prin intermediul unui instrument de alegere a numerelor p şi q astfel ca să fie scrise efectiv în baza cinci (aceasta este scala în care '10' notează numărul *cinci*, '100' îl reprezintă pe *douăzeci şi cinci*, iar '44' pe *douazeci şi patru*, etc.) Dar, decât să folosim simbolurile (cifrele în baza cinci) 0, 1, 2, 3 şi 4, prefer să utilizăm respectivele secvenţe de pe bandă 0, 10, 110, 1110, 11110

Atunci avem următorul tabel de codificare:

0	0	10	1100	20	11110
1	10	11	11010	21	111010
2	110	12	11●110	22	11110110
3	1110	13	1101110	23	11101110
4	11110	14	11011110	24	111011110
5	100	15	11100	25	1000
6	1010	16	111010	26	10010
7	10110	17	1110110	27	100110
8	101110	18	11101110	28	1001110
9	1011110	19	111011110	29	10011110

Notaţia 'C_p' va fi folosită aici pentru specificarea corectă a maşinii Turing T_r, unde r este numărul a cărui expresie binară, cu secvenţa 110 ataşată la sfârşit, este exact expresia lui p în baza cinci, conform schemei de mai sus. Numărul q asupra căruia acţionează calculul C_p, va fi şi el exprimat în baza cinci. Calculul $A(p,q)$ va fi descris ca o maşină Turing acţionând asupra unei benzi codificând perechea de numere p,q. Această codificare pe bandă ar avea forma următoare:

...0011111 0p111110q111110...

unde p şi q sunt reprezentările în baza cinci, ca mai sus, ale lui p şi q.

Ceea ce trebuie este să găsim un p şi un q pentru care să ştim că $C_p(q)$ nu se termină, dar pentru care nici $A(p,q)$ nu se termină. Procedeul din paragraful 2.5 reuşeşte acest lucru găsind un număr k pentru care C_k acţionând asupra lui n este precis $A(n,n)$, pentru fiecare n şi apoi pune $p = q = k$. Pentru a ajunge explicit aici, vom căuta o maşină Turing K (= C_k) a cărei acţiune pe o bandă marcată

...0011111 0m11111000...

este exact aceeaşi cu a lui A asupra:

0011111 0m11111 0m11111000...,

(m fiind reprezentarea în baza cinci, ca mai sus, a lui n), pentru fiecare n.

Atunci, maşina K trebuie să ia numărul n (scris în baza cinci) şi să-l copieze separând cele două apariţii ale lui $\mathbb{m}$ prin secvenţa $\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{0}$ (iar o secvenţă similară se pune la început şi la sfârşit, pentru a marca banda). În rest, K va acţiona asupra benzii transformate exact în acelaşi mod în care ar face-o A cu banda respectivă.

O modificare explicită a lui A care să furnizeze o astfel de maşină K se realizează în felul următor: mai întâi, luăm instrucţiunea iniţială $0\mathbb{1} \to \mathbb{X}$, din specificarea lui A şi analizăm 'conţinutul' lui $\mathbb{X}$. Va trebui să substituim, în cele ce urmează, pe '$\mathbb{X}$' cu respectivul 'conţinut' Ca o chestiune tehnică, trebuie să admitem că A este formulată astfel încât starea internă 0 a lui A nu mai este întâlnită niciodată după execuţia instrucţiunii $0\mathbb{1} \to \mathbb{X}$. Nu există de fapt nici o restricţie asupra lui A în această presupunere [2]

Apoi, trebuie determinat numărul total N de stări interne din specificarea lui A (pornind cu starea 0, starea internă maximă corespunde lui $N-1$). Dacă specificarea lui A nu conţine o instrucţiune finală de forma $(N-1)\mathbb{1} \to \mathbb{Y}$, atunci trebuie adăugată o instrucţiune banală $(N-1)\mathbb{1} \to 0\mathbb{0}\mathbb{R}$ În final, scoatem instrucţiunea $0\mathbb{1} \to \mathbb{X}$ dintre specificaţiile lui A şi adăugăm în schimb instrucţiunile înşirate mai jos, unde fiecare număr de stare internă care apare în listă trebuie majorat cu N, unde θ reprezintă starea finală 0, iar $\mathbb{X}$ în $1\mathbb{1} \to \mathbb{X}$, mai jos, este instrucţiunea discutată anterior (astfel, primele două din instrucţiunile înşirate mai jos vor deveni $0\mathbb{1} \to N\mathbb{1}\mathbb{R}$, respectiv $N\mathbb{0} \to (N+4)\mathbb{0}\mathbb{R}$).

$\theta\mathbb{1} \to 0\mathbb{1}\mathbb{R}$, $0\mathbb{0} \to 4\mathbb{0}\mathbb{R}$, $0\mathbb{1} \to 0\mathbb{1}\mathbb{R}$, $1\mathbb{0} \to 2\mathbb{1}\mathbb{R}$, $1\mathbb{1} \to \mathbb{X}$, $2\mathbb{0} \to 3\mathbb{1}\mathbb{R}$,
$2\mathbb{1} \to \theta\mathbb{0}\mathbb{R}$, $3\mathbb{0} \to 55\mathbb{0}\mathbb{R}$, $3\mathbb{1} \to \theta\mathbb{0}\mathbb{R}$, $4\mathbb{0} \to 4\mathbb{0}\mathbb{R}$, $4\mathbb{1} \to 5\mathbb{1}\mathbb{R}$, $5\mathbb{0} \to 4\mathbb{0}\mathbb{R}$,
$5\mathbb{1} \to 6\mathbb{1}\mathbb{R}$, $6\mathbb{0} \to 4\mathbb{0}\mathbb{R}$, $6\mathbb{1} \to 7\mathbb{1}\mathbb{R}$, $7\mathbb{0} \to 4\mathbb{0}\mathbb{R}$, $7\mathbb{1} \to 8\mathbb{1}\mathbb{R}$, $8\mathbb{0} \to 4\mathbb{0}\mathbb{R}$,
$8\mathbb{1} \to 9\mathbb{1}\mathbb{R}$, $9\mathbb{0} \to 10\mathbb{0}\mathbb{R}$, $9\mathbb{1} \to \theta\mathbb{0}\mathbb{R}$, $10\mathbb{0} \to 11\mathbb{1}\mathbb{R}$, $10\mathbb{1} \to \theta\mathbb{0}\mathbb{R}$,
$11\mathbb{0} \to 12\mathbb{1}\mathbb{R}$, $11\mathbb{1} \to 12\mathbb{0}\mathbb{R}$, $12\mathbb{0} \to 13\mathbb{1}\mathbb{R}$, $12\mathbb{1} \to 13\mathbb{0}\mathbb{R}$, $13\mathbb{0} \to 14\mathbb{1}\mathbb{R}$,
$13\mathbb{1} \to 14\mathbb{0}\mathbb{R}$, $14\mathbb{0} \to 15\mathbb{1}\mathbb{R}$, $14\mathbb{1} \to 1\mathbb{0}\mathbb{R}$, $15\mathbb{0} \to 0\mathbb{0}\mathbb{R}$, $15\mathbb{1} \to \theta\mathbb{0}\mathbb{R}$,
$16\mathbb{0} \to 17\mathbb{0}\mathbb{L}$, $16\mathbb{1} \to 16\mathbb{1}\mathbb{L}$, $17\mathbb{0} \to 17\mathbb{0}\mathbb{L}$, $17\mathbb{1} \to 18\mathbb{1}\mathbb{L}$, $18\mathbb{0} \to 17\mathbb{0}\mathbb{L}$,

---

[2] În fond, una din propunerile originale ale lui Turing a fost ca maşina să se oprească atunci când starea internă '0' este reintrodusă dintr-o altă stare internă. In acest fel, nu numai că restricţia de mai sus nu-i necesară, dar ne putem dispensa de instrucţiunea STOP. O simplificare se va obţine deoarece $1\mathbb{1}\mathbb{1}\mathbb{0}$ nu va mai fi necesar ca instrucţiune şi poate fi folosit ca marcaj, în locul lui $1\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{0}$. Aceasta
va scurta semnificativ prescrierea lui K, putând folosi, în locul codificării în baza cinci, baza patru.

18𝟙 → 19𝟙L, 19𝟘 → 17𝟘L, 19𝟙 → 20𝟙L, 20𝟘 → 17𝟘L, 20𝟙 → 21𝟙L,
21𝟘 → 17𝟘L, 21𝟙 → 22𝟙L, 22𝟘 → 22𝟘L, 22𝟙 → 23𝟙L, 23𝟘 → 22𝟘L,
23𝟙 → 24𝟙L, 24𝟘 → 22𝟘L, 24𝟙 → 25𝟙L, 25𝟘 → 22𝟘L, 25𝟙 → 26𝟙L,
26𝟘 → 22𝟘L, 26𝟙 → 27𝟙L, 27𝟘 → 32𝟙R, 27𝟙 → 28𝟙L, 28𝟘 → 33𝟘R,
28𝟙 → 29𝟙L, 29𝟘 → 33𝟘R, 29𝟙 → 30𝟙L, 30𝟘 → 33𝟘R, 30𝟙 → 31𝟙L,
31𝟘 → 33𝟘R, 31𝟙 → 11𝟘R, 32𝟘 → 34𝟘L, 32𝟙 → 32𝟙R, 33𝟘 → 35𝟘L,
33𝟙 → 33𝟙R, 34𝟘 → 36𝟘R, 34𝟙 → 34𝟘R, 35𝟘 → 37𝟙R, 35𝟙 → 35𝟘R,
36𝟘 → 36𝟘R, 36𝟙 → 38𝟙R, 37𝟘 → 37𝟘R, 37𝟙 → 39𝟙R, 38𝟘 → 36𝟘R,
38𝟙 → 40𝟙R, 39𝟘 → 37𝟘R, 39𝟙 → 41𝟙R, 40𝟘 → 36𝟘R, 40𝟙 → 42𝟙R,
41𝟘 → 37𝟘R, 41𝟙 → 43𝟙R, 42𝟘 → 36𝟘R, 42𝟙 → 44𝟙R, 43𝟘 → 37𝟘R,
43𝟙 → 45𝟙R, 44𝟘 → 36𝟘R, 44𝟙 → 46𝟙R, 45𝟘 → 37𝟘R, 45𝟙 → 47𝟙R,
46𝟘 → 48𝟘R, 46𝟙 → 46𝟙R, 47𝟘 → 49𝟘R, 47𝟙 → 47𝟙R, 48𝟘 → 48𝟘R,
48𝟙 → 49𝟘R, 49𝟘 → 48𝟙R, 49𝟙 → 50𝟙R, 50𝟘 → 48𝟙R, 50𝟙 → 51𝟙R,
51𝟘 → 48𝟙R, 51𝟙 → 52𝟙R, 52𝟘 → 48𝟙R, 52𝟙 → 53𝟙R, 53𝟘 → 54𝟙R,
53𝟙 → 53𝟙R, 54𝟘 → 16𝟘L, 54𝟙 → 𝜃𝟘R, 55𝟘 → 53𝟙R

Acum suntem capabili să găsim limita precisă a dimensiunii lui K, aşa cum rezultă din construcţia de mai sus, ca o funcţie de dimensiunea lui A. Să măsurăm această 'dimensiune' prin 'gradul de complicaţie', definit în paragraful 2.6 (sfârşitul răspunsului la **Q8**). Pentru o maşină Turing anume, T_m (aşa ca A), acesta este numărul de cifre din reprezentarea binară a numărului m. Pentru acţiunea unei maşini Turing *action* $T_m(n)$ (aşa cum este K), este vorba de numărul cifrelor binare din reprezentarea celui mai mare dintre m şi n Fie α şi κ numărul cifrelor binare din a şi, respectiv, k', unde $A = T_a$ şi $K = T_{k'}$ ($= C_k$).

Cum A are cel puţin $2N$-1 instrucţiuni (prima find omisă), şi cum fiecare instrucţiune conţine cel puţin 3 cifre binare în specificarea sa binară, numărul total de cifre binare din numărul a al maşinii Turing trebuie neapărat să satisfacă $\alpha \geq 6N-6$. Există 105 locuri (în dreapta săgeţilor) în lista adiţională de mai sus cu instrucţiuni a lui K, unde numărul N trebuie adunat la numărul ce apare acolo. Numerele rezultate sunt toate mai mici decât $N+55$ şi deci au reprezentări binare expandate cu nu mai mult decât $2\log_2(N+55)$ cifre, ceea ce înseamnă în total mai puţin decât $210\log_2(N+55)$ cifre binare pentru specificările suplimentare de stări interne. Trebuie să adunăm la asta cifrele necesare pentru simbolurile suplimentare, 𝟘, 𝟙, **R** şi **L**, ceea ce măreşte cu alte 527 (incluzând o posibilă

instrucțiune banală și având în minte faptul că 6 din 0-uri pot fi eliminate în concordanță cu regula prin care 00 poate fi reprezentat ca 0), astfel că putem fi siguri că specificarea lui K necesită mai puțin cu $527 + 210\log_2(N+55)$ cifre decât are A, adică:

$$\kappa < \alpha + 527 + 210\log_2(N+55)$$

Folosind relația $\alpha \geq 6N-6$ obținută mai sus, găsim (știind că $210\log_2 6 > 542$)

$$\kappa < \alpha - 15 + 210\log_2(\alpha+336).$$

Acum, să găsim gradul de complicație η al unui calcul particular $C_k(k)$ dat de această procedură. Amintim că gradul de complicație a lui $T_m(n)$ a fost definit ca numărul cifrelor binare din cel mai mare dintre m și n. În situația prezentă, avem $C_k = T_{k'}$, astfel că numărul de cifre binare din 'm' pentru acest calcul este exact κ. Pentru a vedea câte cifre binare sunt în 'n' pentru acest calcul, examinăm banda ce este folosită în $C_k(k)$. Această bandă este inițiată de o secvență 111110, care e imediat urmată de expresia binară a lui k', și apoi se termină cu secvența 11011111. Convențiile din **MN** cer ca această secvență întreagă, dar cu ultima cifră ștearsă, să fie citită ca un număr binar, pentru a obține 'n'-ul care numără banda în calculul $T_m(n)$. Prin urmare, numărul de cifre binare al acestui 'n' particular este precis $\kappa+13$, astfel că rezultă că $\kappa+13$ este gradul de complicație η al lui $C_k(k)$, căci $\eta = \kappa + 13 < \alpha - 2 + 210\log_2(\alpha+336)$, de unde urmează expresia mai simplă:

$$\eta < \alpha + 210\log_2(\alpha+336)$$

Detaliile raționamentului precedent sunt elemente ale codificărilor particulare adoptate pentru mașinile Turing și acestea pot fi diferite în alte codificări. Ideea de bază este una foarte simplă. În esență, este adoptat un formalism de tip λ-*calcul*, apoi întreaga operație va fi, într-un anumit sens, redusă până la trivialitate (vezi **MN**, sfârșitul capitolului 2, pentru o descriere adecvată a λ-calculului lui Church; vezi și Church 1941). Ne putem gândi la A ca fiind definit printr-un operator **A** de λ-calcul, acționând asupra altor operatori **P** și **Q**, așa cum este exprimat de **(AP)A** Aici **P** reprezintă calculul C_p, și **Q** corespunde numărului q. Atunci condiția pentru **A**, pentru orice astfel de **P, Q**, ar fi:

„Dacă **(AP)Q** se termină, atunci **PQ** nu se termină"

Se poate, destul de simplu, construi o operație de λ-calcul care nu se termină, dar pentru care **A** nu îndeplinește această condiție. Asta se obține luând

$\mathbf{K} = \lambda x.[(\mathbf{A}x)x]$ astfel încât $\mathbf{KY} = (\mathbf{AY})\mathbf{Y}$ pentru orice operator $\mathbf{Y}$ Atunci considerăm operaţia de λ-calcul $\mathbf{KK}$ Este limpede că aceasta nu se termină, deoarece $\mathbf{KK} = (\mathbf{AK})\mathbf{K}$, terminarea căreia ar implica că $\mathbf{KK}$ nu se termină, prin natura presupusă a lui $\mathbf{A}$. Mai mult, $\mathbf{A}$ nu poate determina acest fapt deoarece $(\mathbf{AK})\mathbf{K}$ nu se termină. Dacă vom crede că $\mathbf{A}$ are proprietatea cerută, atunci trebuie să credem că $\mathbf{KK}$ nu se termină.

Notăm că există o considerabilă economie în această schemă. Dacă scriem $\mathbf{KK}$ în forma $\mathbf{KK} = \lambda y.[yy](\lambda y.[(\mathbf{A}x)x])$, atunci vedem că numărul de simboluri din $\mathbf{KK}$ este exact de 16 ori mai mare decât numărul din A (ignorând punctele, care sunt redondante, în orice caz!).

Vorbind strict, aceasta nu-i pe de-a-întregul legitim, deoarece simbolul 'x' poate să apară şi în expresia pentru $\mathbf{A}$ şi deci vor trebui adoptate anumite metode de tratare a acestei situaţii. În acest calcul, care nu se termină, este dificil de acceptat faptul că ceeace generează această procedură nu apare ca operaţie asupra numerelor naturale (cât timp al doilea $\mathbf{K}$ în $\mathbf{KK}$ nu este 'număr'). În fond, λ-calculul nu este foarte potrivit pentru operaţii numerice explicite şi nu este prea uşor să vezi cum o procedură algoritmică dată, aplicată numerelor naturale, poate fi exprimată ca o operaţie de λ-calcul. Pentru motive precum acestea, discuţia în termenii maşinilor Turing are o mai mare semnificaţie pentru scopurile noastre căci ceea ce se cere se obţine mai simplu.

3. DESPRE NON-COMUTABILITATE [1] ÎN GÂNDIREA MATEMATICĂ

3.1. Ce au gândit Gödel şi Turing ?

În capitolul 2 am încercat să conving cititorul de natura riguroasă şi de substratul puternic al afirmaţiei (notată cu $\mathcal{G}$), conform căreia înţelegerea nu poate fi rezultatul unui algoritm apreciat conştient şi în totalitate credibil (sau a mai multor algoritmi; vezi **Q1**). Argumentele mele nu se referă la posibilitatea mai gravă ca — *în acord* cu $\mathcal{G}$ — credibilitatea matematică să fie rezultatul unui algoritm necunoscut şi inconştient sau, eventual, al unui algoritm cognoscibil despre care însă să nu se ştie — sau să nu se creadă cu fermitate — că este cel care stă la baza credibilităţii matematice. Scopul meu, în acest capitol, este să arăt că, deşi acestea ar fi posibilităţi logice, ele nu sunt deloc plauzibile.

Mai întâi trebuie subliniat că, atunci când matematicienii produc lanţul riguros al unui raţionament cu care stabilesc adevărul matematic, ei nu se gândesc că urmează pur şi simplu orbeşte reguli inconştiente pe care nu sunt capabili nici să le cunoască şi nici să le creadă. Ceea ce gândesc este că atunci când construiesc

[1] Termenul *non-computabilitate* îl vom folosi pentru a desemna imposibilitatea modelării prin calcule efectuate de un calculator, spre deosebire de *necalculabilitate*, care se referă la caracterul necalculabil, în principiu. (n.t)

un raţionament, îşi bazează argumentele pe ceva ce este fără îndoială adevărat -- şi, în cele din urmă în mod esenţial *evident*. Cu toate că, uneori, acest lanţ poate să fie foarte lung, sau conceptual subtil, raţionamentul este, în principiu, indubitabil, credibil şi logic impecabil. Ei nu sunt de părere că ar acţiona în concordanţă cu anumite proceduri necunoscute sau de neimaginat, care ar sta „în spatele scenei", manipulându-le convingerile în moduri ce nu pot fi cunoscute.

Evident, matematicienii ar putea să se înşele în acest caz. Într-adevăr, poate că există o procedură algoritmică care, necunoscută de ei, le guvernează toate percepţiile matematice. Probabil, este mai uşor pentru ne-matematicieni să ia o astfel de posibilitate în serios, decât este pentru cei mai mulţi dintre matematicienii de profesie. În acest capitol, voi încerca să-l conving pe cititor că aceşti matematicieni profesionişti au dreptate în părerea lor că ei nu lucrează pur şi simplu sub comanda unui algoritm necunoscut (şi incognoscibil) şi nici a unui algoritm în care nu s-ar încrede cu fermitate. S-ar putea foarte bine ca gândurile şi convingerile lor să fie într-adevăr ghidate de nişte principii inconştiente; dar eu susţin că dacă este aşa, atunci acele principii sunt lucruri care nu pot fi descrise în termeni algoritmici.

E instructiv să prezentăm punctele de vedere ale celor două personalităţi care sunt, esenţial, responsabile pentru rezultatul ce ne-a condus la concluzia noastră $\mathcal{G}$

De fapt, ce gândea Gödel? Ce gândea Turing? Este interesant de văzut cum propunând aceeaşi evidenţă matematică, ei au ajuns la concluzii diametral opuse. Trebuie însă să fie clar că amândoi aceşti gânditori de excepţie au exprimat puncte de vedere în concordanţă cu concluzia $\mathcal{G}$

Punctul de vedere a lui Gödel pare să fi fost acela că nu există constrângere care să facă din minte o entitate computaţională, iar pe de altă parte, mintea nu este constrânsă de finitudinea creierului. Într-adevăr, el l-a certat pe Turing pentru că nu accepta această posibilitate.

După cum spune Hao Wang (1974, p.326 şi conform cu Gödel, Collected Works vol II, p.297, 1990), în timp ce accepta ambele afirmaţii controversate ale lui Turing „creierul funcţionează ca un calculator" şi „legile fizicii, în consecinţele lor observabile, au o limită finită de precizie", Gödel a respins o altă afirmaţie controversată a lui Turing şi anume că „nu există minte separată de materie", referindu-se la ea ca la „un prejudiciu adus vremurilor noastre"

Astfel, se vede că Gödel lua drept evident faptul că creierul *fizic* poate să se comporte în conformitate cu legile calculului, dar că mintea este ceva deasupra creierului, astfel încât acţiunea minţii nu este constrânsă să se comporte după legile calculului, despre care el credea că trebuie să comande comportarea fizică a creierului. El nu privea afirmaţia $\mathcal{G}$ ca pe o demonstraţie a punctului său de vedere, că mintea ar acţiona non-computaţional, deoarece el admitea [1], de fapt, că:

> „Pe de altă parte, pe baza a ceea ce s-a demonstrat până aici, rămâne posibil ca să existe (ba chiar să poată fi descoperită empiric) o maşină de demonstrat teoreme, care de fapt *este* echivalentă cu intuiţia matematică, dar despre care nu se poate *demonstra* că este aşa, nici chiar că ea produce numai teoreme *corecte* din teoria finitară a numerelor."

Ar trebui să fie clar că această idee nu contravine încrederii în $\mathcal{G}$ (şi eu nu am nici o îndoială că Gödel era în posesia unei concluzii tot atât de clare ca şi formularea pe care am dat-o eu în $\mathcal{G}$). El a acceptat *posibilitatea logică* că mintea matematicienilor ar putea funcţiona în conformitate cu un algoritm de care ei să nu fie conştienţi, sau poate de care ar putea fi conştienţi, dar de a cărui corectitudine ar putea să nu fie absolut convinşi (" nu se poate demonstra că produce numai teoreme *corecte* "). Un astfel de algoritm intră în categoria algoritm „necognoscibil corect", în terminologia mea, adică a cărui corectitudine nu poate fi cunoscută. Desigur, o altă problemă ar fi să credem că un astfel de algoritm, despre care să nu se ştie că este corect, ar putea să stea la baza activităţii mentale a unui matematician. Gödel nu a crezut aşa ceva, ci s-a simţit condus în direcţia mistică pe care eu am marcat-o cu $\mathcal{D}$ — în conformitate cu care, mintea nu poate fi explicată în termenii ştiinţei despre lumea fizică.

Pe de altă parte, Turing a respins un astfel de punct de vedere, crezând (ca şi Gödel) că, la fel cu oricare alt obiect fizic, creierul trebuie să acţioneze computaţional (să ne reamintim teza lui Turing din paragraful 1.6). Astfel, Turing trebuia să găsească alte căi privind problema pe care o ridică $\mathcal{G}$

Turing aprecia ca foarte important faptul că matematicienii pot foarte bine să facă greşeli; şi argumenta că pentru ca un calculator să fie cu adevărat inteligent, el ar trebui de asemenea să poată face greşeli [2]:

„Cu alte cuvinte, dacă despre o maşină se aşteaptă să fie infailibilă, atunci ea nu poate fi şi inteligentă. Există câteva teoreme care spun aproape acest lucru. Dar aceste teoreme nu spun nimic despre câtă inteligenţă poate afişa maşina dacă nu are nici o pretenţie de infailibilitate."

Teoremele pe care le avea Turing în minte sunt, fără îndoială, teorema lui Gödel şi alte câteva legate de ea, aşa cum este chiar propria sa versiune *computaţională* a teoremei lui Gödel. Astfel, el priveşte imprecizia gândirii matematice a oamenilor, ca esenţială, permiţând minţii, (presupus) imprecisă algoritmic, să furnizeze o capacitate mai mare decât cea care ar putea fi atinsă cu ajutorul oricărei proceduri algoritmice corecte. În acest mod, el sugerează o posibilitate de a evada din concluziile teoremei lui Gödel: algoritmul matematicianului ar fi tehnic incorect şi, cu siguranţă, nu ar fi 'cunoscut drept corect'. Astfel, punctul de vedere a lui Turing este în concordanţă cu $\mathcal{G}$ şi pare probabil că el ar fi fost de acord cu poziţia descrisă de $\mathcal{A}$.

Ca parte a discuţiei care urmează, voi prezenta motivele pentru care eu nu cred că 'incorectitudinea' din algoritmul matematicianului ar putea fi, eventual, considerată explicaţia *reală* a ceea ce se petrece în mintea matematicianului. Există, în orice caz, ceva inacceptabil în ideea că ceea ce face ca o minte să fie superioară unui calculator este *in*adecvarea în special atunci când, ca aici, este vorba nu despre originalitatea sau creativitatea matematicienilor, ci pur şi simplu, despre capacitatea lor de a percepe *adevărul matematic indubitabil*. Este uimitor faptul că fiecare dintre aceşti doi mari gânditori, Gödel şi Turing, au fost conduşi, din consideraţii ca $\mathcal{G}$, într-o poziţie pe care mulţi ar considera-o ca fiind de neacceptat. Este interesant să ne întrebăm cum de au ajuns în poziţia de a contempla posibilitatea serioasă ca acţiunea fizică să fie, la origine, ceva non-computaţional —în concordanţă cu punctul de vedere $\mathcal{C}$, pe care eu îl promovez aici.

În secţiunile următoare (în special în paragrafele 3.2-3.22) voi prezenta detaliat câteva argumente, unele dintre ele destul de complicate, confuze sau tehniciste care au ca scop să spulbere atât modelul computaţional $\mathcal{A}$, cât şi $\mathcal{B}$, care ar furniza o bază plauzibilă pentru înţelegerea matematică. Recomandarea mea pentru cititorul care nu are nevoie să fie convins —sau care-i îngrozit de aceste detalii este să citească atât cât este dispus, iar atunci, când totul va părea prea anost, să treacă direct la dialogul fantastic (din paragraful 3.23) care prezintă un

rezumat al ideilor expuse. Vă rog să vă întoarceţi la cele de mai jos, numai dacă şi când, veţi simţi nevoia.

3.2. Poate un algoritm incorect să simuleze înţelegerea matematică într-un mod care poate fi cunoscut?

Trebuie să considerăm, în concordanţă cu $\mathcal{G}$, că înţelegerea matematică ar putea fi rezultatul unui algoritm incorect sau care nu poate fi cunoscut, sau a unui algoritm corect şi cognoscibil dar despre care nu se poate demonstra că este corect.

Prin *algoritm* înţelegem doar o anumită procedură de calcul (ca în paragraful 1.5) — adică orice lucru care poate fi simulat principial de un calculator cu o memorie nelimitată (trebuie să reamintesc, din discuţia privind **Q8** din paragraful 2.6, că natura *nelimitată* a memoriei, în această idealizare, nu slăbeşte argumentaţia). În continuare, această noţiune de *algoritm* va include procedurile *top-down* şi sistemele de învăţare *bottom-up*, precum şi combinaţii ale acestora. Va fi luat în considerare, de exemplu, tot ceea ce poate fi obţinut prin reţele neuronale (vezi paragraful 1.5). La fel, vor fi incluse şi alte tipuri *bottom-up* de mecanisme, ca cele numite *algoritmi genetici*, care pot să se autoperfecţioneze cu ajutorul unui algoritm gata construit, prin analogie cu teoria darwinistă a evoluţiei (vezi paragraful 3.11).

În paragrafele 3.9—3.22 (rezumate în dialogul fantastic din paragraful 3.23) voi arăta efectiv cum procedurile *bottom-up* sunt importante pentru justificarea pe care o voi da în acest paragraf. Totuşi, pentru claritate, momentan mă voi limita la algoritmii *top-down*. Această acţiune algoritmică poate fi considerată relevantă, atât pentru un matematician anume, cât şi pentru comunitatea matematicienilor ca un întreg.

În discutarea întrebărilor **Q11** şi **Q12**, din paragraful 2.10, am considerat posibilitatea ca să existe *diferiţi* algoritmi corecţi şi cunoscuţi acceptabili pentru persoane diferite şi am ajuns la concluzia că această posibilitate nu afectează raţionamentul. Vom examina ceva mai târziu (în paragraful 3.7) posibilitatea ca să existe (să fie utilizaţi) algoritmi incorecţi sau care nu pot fi cunoscuţi.

Deocamdată, să presupunem că ar exista o singură procedură algoritmică care stă la baza înţelegerii matematice. Ne vom restrânge de asemenea atenţia la acele elemente ale înţelegerii matematice care se utilizează pentru a demonstra propoziţii π_1 (adică, pentru specificarea unei maşini Turing a cărei acţiune nu se termină niciodată, conform cu **Q10**). În cele ce urmează, este suficient dacă vom interpreta expresia „înţelegere matematică" în acest sens restrictiv (vezi $\mathcal{G}^{**}$ paragraful 2.10.).

Trebuie să facem diferenţa între cele trei variante distincte cu privire la posibilitatea de a cunoaşte o presupusă procedură algoritmică **F**, care ar sta la baza înţelegerii matematice, procedură corectă sau nu. Astfel, **F** ar putea fi:

I o procedură care poate fi cunoscută în mod conştient şi al cărei rol de algoritm al înţelegerii matematice poate fi şi el cunoscut;

II o procedură care poate fi cunoscută în mod conştient, dar al cărei rol, ca algoritm care stă la baza înţelegerii matematice este inconştient şi nu poate fi cunoscut;

III o procedură inconştientă şi necognoscibilă.

Să considerăm mai întâi cazul complet conştient **I**. Deoarece atât algoritmul cât şi rolul său pot fi cunoscute, putem presupune foarte bine că ele sunt *deja* cunoscute. În imaginaţia noastră, putem presupune că ne plasăm cu argumentaţia într-o vreme în care acestea sunt cu adevărat cunoscute — deoarece *cognoscibil* înseamnă că, cel puţin în principiu, un astfel de moment va veni. Deci, să presupunem că algoritmul **F** este unul cunoscut, rolul său fiind de asemenea cunoscut. Am văzut în paragraful 2.9 că un astfel de algoritm este efectiv echivalent cu un sistem formal $\mathbb{F}$. Astfel, am presupus că înţelegerea matematică — sau cel puţin aceea a unui anumit matematician — este cunoscută (de către acel matematician) ca fiind echivalentă cu posibilitatea de derivabilitate într-un anumit sistem formal $\mathbb{F}$. Pentru a avea o speranţă să satisfacem concluzia noastră $\mathcal{G}$, trebuie să presupunem că sistemul nostru $\mathbb{F}$ nu este corect. În mod destul de surprinzător, incorectitudinea sistemului nu ne ajută deloc în cazul unui sistem formal cunoscut $\mathbb{F}$ care, aşa cum am presupus în **I**, este în mod actual cunoscut — şi astfel crezut de către orice matematician că stă la baza înţelegerii sale matematice! Pentru că o astfel de convingere îndreptăţeşte încrederea (greşită) în

corectitudinea lui $\mathbb{F}$ (Ar fi o poziţie matematică neraţională aceea care ar permite neîncrederea în chiar baza propriului sistem indubitabil de convingeri!)

Indiferent dacă $\mathbb{F}$ este sau nu corect, încrederea în faptul că este de fapt corect, îndreptăţeşte convingerea că $G(\mathbb{F})$ (sau cu altă notaţie $\mathcal{Q}(\mathbb{F})$, ca în paragraful 2.8) este adevărată; dar cum $G(\mathbb{F})$ pe baza acceptării teoremei lui Gödel – este considerat că s-ar situa în afara domeniului lui $\mathbb{F}$ acest lucru contrazice presupunerea că $\mathbb{F}$ stă la baza *tuturor* înţelegerilor matematice (relevante). Argumentaţia merge tot atât de bine când se referă la un matematician anume sau la comunitatea matematicienilor; ea se poate aplica separat tuturor algoritmilor care ar putea fi presupuşi a sta la baza procesului de gândire al matematicienilor. Mai mult, trebuie să se refere doar la acea înţelegere matematică care permite stabilirea propoziţiilor π_1. Astfel, oricare ar fi un aşa-zis algoritm $\mathbb{F}$ despre care se ştie că e corect şi despre care s-ar presupune că ar sta la baza înţelegerii matematice, algoritmul $\mathbb{F}$ nu poate fi *recunoscut* că ar fi astfel, atâta vreme cât **I** este respinsă, indiferent dacă $\mathbb{F}$ este sau nu un sistem corect. Dacă însă $\mathbb{F}$ nu poate fi cunoscut, atunci suntem în cazul **II**: şi anume, $\mathbb{F}$ ar putea, de fapt, să stea la baza înţelegerii matematice, dar, acest rol al său nu poate fi cunoscut. Mai există posibilitatea **III** ca sistemul $\mathbb{F}$ să fie el însuşi inconştient şi incognoscibil.

Ceea ce am stabilit, în esenţă, până aici, este că varianta **I** (cel puţin în contextul algoritmilor *top-down*) nu este o posibilitate serioasă; faptul că sistemul $\mathbb{F}$ ar putea să nu fie corect nu are, de fapt, nici o importanţă pentru **I**. Lucrul cu adevărat important este că $\mathbb{F}$, corect sau nu, nu poate fi cunoscut ca fiind la baza încrederii matematice; nu incognoscibilitatea algoritmului în sine este problema, ci că nu poate fi cunoscut faptul că acest algoritm *este* cu adevărat baza înţelegerii.

3.3. Poate simula înţelegerea matematică un algoritm cognoscibil într-un mod care nu poate fi cunoscut?

Vom trata acum cazul **II** şi vom încerca să luăm în serios posibilitatea ca înţelegerea matematică să fie de fapt echivalentă cu un algoritm sau un sistem formal care poate fi cunoscut în mod conştient, dar a cărui funcţionare nu poate fi

cunoscută. Astfel, deşi sistemul presupus, $\mathbb{F}$, ar fi unul cognoscibil, noi n-am putea fi niciodată siguri că *acest* sistem particular stă la baza înţelegerii matematice. Să vedem dacă această posibilitate este plauzibilă.

Dacă presupusul $\mathbb{F}$ nu este un sistem formal deja cunoscut, atunci, ca mai înainte, trebuie să considerăm că, într-o bună zi el va fi cunoscut, cel puţin în principiu. Să ne imaginăm că acea zi a sosit şi că specificarea precisă a lui $\mathbb{F}$ se află în faţa noastră. Deşi s-ar putea ca acest sistem formal $\mathbb{F}$ să fie foarte elaborat, presupunem că este destul de simplu pentru ca noi să fim în stare, cel puţin în principiu, să-l putem, într-un mod foarte conştient, aprecia dar nouă nu ni se permite să fim siguri că $\mathbb{F}$ conţine de fapt cu siguranţă totalitatea lucrurilor care pot fi indubitabil înţelese matematic şi a intuiţiilor matematice (cel puţin cele relative la propoziţiile π_1). Vom încerca să vedem de ce acest lucru, deşi logic este posibil, este de fapt foarte puţin plauzibil. Mai mult, voi încerca să arăt în continuare că, chiar dacă ar fi adevărată, o astfel de posibilitate nu ar oferi nici o satisfacţie practicienilor din **IA** care ar intenţiona să construiască un robot matematician!. Acest aspect al lucrurilor va fi reluat la sfârşitul acestui paragraf şi va fi dezvoltat mai complet în 3.15 şi 3.29.

Pentru a sublinia faptul că existenţa unui astfel de $\mathbb{F}$ trebuie considerată ca o posibilitate logică, trebuie să reamintim „maşina de demonstrat teoreme" despre care Gödel remarca că nu poate fi (încă) respinsă logic (aşa cum rezultă din citatul dat în paragraful 3.1) De fapt, această 'maşină' ar trebui, aşa cum voi explica, să fie o procedură algoritmică **F** care să fie în concordanţă cu unul dintre cazurile **II** sau **III**. Presupusa maşină de demonstrat a lui Gödel, aşa cum spunea el, ar putea fi 'empiric descoperită', lucru ce corespunde cerinţei ca **F** să poată fi 'cunoscut conştient' ca în **II**, sau să nu fie, caz ce corespunde condiţiilor din **III**.

Gödel a argumentat, pe baza teoremei sale, că procedura **F** (sau echivalent, sistemul formal $\mathbb{F}$; conform cu 2.9) nu poate fi 'demonstrată' ca „echivalentă cu intuiţia matematică" – după cum susţin cuvintele citate. În **II** (şi prin implicaţie, în **III**, de asemenea) am formulat acestă limitare fundamentală a lui **F** într-un mod oarecum diferit: „rolul acestui algoritm ca bază a înţelegerii matematice este inconştient şi nu poate fi cunoscut"

Această limitare (a cărei necesitate rezultă din respingerea lui **I**, aşa cum am argumentat-o în paragraful precedent), în mod clar nu implică faptul că despre F nu s-ar putea arăta că este echivalent cu intuiţia matematică, deoarece o astfel de

demonstrație ar stabili, pentru noi, că **F** ar avea chiar rolul pe care am presupus că nu se poate cunoaște că îl are.

Invers, dacă chiar acest rol al lui **F** în a furniza baza de necontestat pentru înțelegerea matematică, ar fi ceva care ar putea fi cunoscut conștient — în sensul că s-ar putea aprecia pe deplin cum acționează **F** în acest rol — atunci corectitudinea lui **F** ar trebui acceptată de asemenea. Pentru că dacă **F** nu ar fi acceptat pe deplin drept corect, acest lucru ar conduce la refuzul de a accepta unele dintre consecințele sale. Dar consecințele sale sunt exact acele propoziții matematice (sau cel puțin propozițiile π_1) care ar trebui acceptate. Astfel, cunoașterea rolului lui **F** ne pune în posesia unei *demonstrații* a lui **F**, deși această demonstrație nu ar fi una formală, într-un sistem formal predefinit.

Vom observa de asemenea că propozițiile π_1 valide pot fi privite ca exemple de „teoreme corecte ale teoriei finitare a numerelor" la care se referă Gödel. De fapt, dacă termenul „teoria finitară a numerelor" ar fi considerat că include operația μ „să se găsească cel mai mic număr natural cu proprietatea aritmetică cutare sau cutare", caz în care ea ar include acțiunile unei mașini Turing (vezi paragraful 2.8), atunci *toate* propozițiile π_1 ar conta ca făcând parte din teoria finitară a numerelor. Astfel, ar rezulta că raționamentul de tip Gödel nu ar furniza o metodă clară de respingere a cazului **II** numai pe baze logice riguroase — cel puțin dacă acceptăm autoritatea lui Gödel!

Pe de altă parte, trebuie să ne întrebăm dacă **II** poate fi, în vreun fel, o posibilitate plauzibilă. Să încercăm să vedem dacă se impune existența unui sistem logic $\mathbb{F}$ care deși poate fi cunoscut, nu se poate ști că este echivalent cu înțelegerea matematică (indubitabilă). Așa cum remarcam mai sus, ne putem imagina că a venit timpul când acest $\mathbb{F}$ a fost descris și se află în fața noastră. Să ne reamintim (vezi paragraful 2.7) că un sistem formal este specificat în termenii unei mulțimi de axiome și reguli de procedură. Teoremele din $\mathbb{F}$ sunt propoziții care pot fi obținute din axiome cu ajutorul regulilor de procedură, toate teoremele fiind lucruri ce pot fi formulate în termenii aceloraşi simboluri cu care au fost exprimate axiomele. Ceea ce încercăm să ne imaginăm este că teoremele lui $\mathbb{F}$ sunt exact acele propoziții (scrise cu acele simboluri) care pot, în principiu, să fie percepute ca sigur adevărate, de către matematicienii oameni.

Să presupunem, pentru moment, că lista axiomelor ar fi finită. Dar, axiomele însele pot conta totdeauna ca instanțieri particulare ale teoremelor. Căci

fiecare teoremă este ceva care poate, în principiu, să fie perceput ca fără îndoială adevărat, utilizând înţelegerea umană şi intuiţia. Astfel, fiecare axiomă luată individual trebuie să exprime ceva care, în principiu, este o parte a acestei înţelegeri matematice. În acest fel, pentru fiecare axiomă individuală, vine o vreme (sau în orice caz, în principiu, poate veni o astfel de vreme) când este percepută ca fiind adevărată fără nici o îndoială. Una după alta, axiomele individuale ale lui $\mathbb{F}$ ar putea fi percepute ca atare. Astfel *toate* axiomele individuale ale lui $\mathbb{F}$ vor fi percepute ca adevărate fără îndoială (sau ar putea fi, în principiu, astfel percepute). Ca urmare, ar putea să vină o vreme când totalitatea axiomelor lui $\mathbb{F}$, ca întreg, să fie considerate ca valide, fără vreo îndoială.

Ce se poate spune despre regulile de procedură? Putem întrezări o vreme când acestea ar fi considerate corecte, cu siguranţă? Pentru multe sisteme formale aceste lucruri sunt simple şi toate sunt fără îndoială acceptabile, ca de exemplu: „Dacă am stabilit deja că **P** este o teoremă şi de asemenea că $\mathbf{P}\Rightarrow\mathbf{Q}$ este o teoremă, atunci putem deduce că **Q** este o nouă teoremă" (vezi **MN**, sau Kleene 1952, cu privire la simbolul *implicaţie* $\Rightarrow$). Pentru astfel de reguli, nu vor fi nici un fel de dificultăţi să fie acceptate ca de necontestat. Pe de altă parte, ar putea să existe mijoace de inferenţă mult mai subtile, incluse printre regulile de procedură, a căror validitate să nu fie chiar atât de evidentă şi care să necesite consideraţii delicate înainte de a ajunge la o decizie dacă să se accepte sau nu astfel de reguli ca „fără indoială corecte" De fapt, aşa cum vom vedea îndată, pot să existe printre regulile de procedură ale lui $\mathbb{F}$, unele a căror corectitudine neîndoielnică să nu poată fi percepută de matematicieni — chiar dacă am presupune că numărul axiomelor lui $\mathbb{F}$ este finit.

De ce am înşirat toate acestea? Să revenim, în imaginaţia noastră la vremea în care toate axiomele sunt considerate cu necesitate adevărate. Să ne aşezăm şi să contemplăm sistemul $\mathbb{F}$. Să încercăm să presupunem că regulile de procedură ale lui $\mathbb{F}$ sunt şi ele lucruri pe care acum le putem accepta ca de necontestat. Deşi nu se presupune că am fi în stare să ştim că $\mathbb{F}$ conţine chiar toată matematica care este, în principiu, accesibilă înţelegerii umane şi intuiţiei, de aici înainte, noi, cel puţin, ne-am convins că $\mathbb{F}$ este sigur corect, deoarece atât axiomele sale cât şi regulile de procedură sunt acceptate ca adevărate. Acum noi ar trebui să fim efectiv convinşi că $\mathbb{F}$ este *consistent*. Ne-ar putea veni totuşi o idee şi anume că $G(\mathbb{F})$ ar trebui să fie de asemenea adevărată, în virtutea acestei consistenţe — adevărată fără

îndoială! Dar cum $\mathbb{F}$ este presupus, de fapt (dar de neaflat pentru noi) că conţine totalitatea a ceea ce ne este cu siguranţă accesibil nouă, $G(\mathbb{F})$ trebuie să fie de fapt o teoremă în $\mathbb{F}$. Dar, din teorema lui Gödel, acest lucru s-ar putea întâmpla numai dacă $\mathbb{F}$ ar fi de fapt inconsistent. Dacă $\mathbb{F}$ este inconsistent, atunci '1=2' este o teoremă în $\mathbb{F}$. Astfel afirmaţia 1=2 ar face parte, în principiu, din înţelegerea noastră matematică de neatacat – ceea ce este cu siguranţă o contradicţie!

Cu toate acestea, susţinem, cel puţin ca *posibilitate*, că oamenii care se ocupă de matematică ar putea opera, fără să ştie, în concordanţă cu $\mathbb{F}$, care nu este de fapt corect. Las această chestiune pentru paragraful 3.4; dar, pentru nevoile acestei secţiuni, să acceptăm că procedurile care stau la baza înţelegerii matematice sunt de fapt perfect corecte. În astfel de circumstanţe, am ajuns la o contradicţie, dacă ar fi să presupunem că regulile de procedură ale sistemului nostru $\mathbb{F}$, cu număr finit de axiome are toate axiomele acceptabile, fără rezerve. Astfel, ar trebui să se afle, printre regulile de procedură ale lui $\mathbb{F}$, cel puţin una care să nu poată fi percepută de către un matematician profesionist, ca sigur corectă (deşi ea ar putea fi corectă).

Toate acestea au loc când $\mathbb{F}$ are un număr finit de axiome. O posibilă generalizare ar fi să presupunem că $\mathbb{F}$ ar putea avea un număr *infinit* de axiome. Voi comenta această posibilitate. Pentru ca sistemul $\mathbb{F}$ să poată fi calificat drept sistem formal în sensul cerut – adică astfel încât totdeauna să se poată verifica, cu ajutorul unei proceduri predefinite, faptul că o demonstraţie a unei anumite propoziţii este cu adevărat o demonstraţie conformă cu regulile lui $\mathbb{F}$ – este necesar ca sistemul său infinit de axiome să poată fi exprimat în termeni finiţi. În fapt, există totdeauna o oarecare libertate în ceea ce priveşte modul în care un anumit sistem formal este reprezentat, prin care operaţiile sale sunt desemnate ca axiome sau reguli de procedură. De fapt, sistemul axiomatic standard pentru teoria mulţimilor – sistemul Zermelo-Fraenkel (notat cu $\mathbb{ZF}$) – posedă o infinitate de axiome, exprimate în termenii unei structuri numită *schemă axiomatică*. Printr-o reformulare corespunzătoare, sistemul $\mathbb{ZF}$ poate fi exprimat astfel încât numărul de axiome să devină finit [3]. De fapt, într-un anumit sens, acest lucru poate fi făcut întotdeauna pentru sistemele axiomatice care sunt sisteme formale în sensul cerut aici[2], al calculabilităţii.

Există o variantă banală în care acest lucru poate fi făcut şi anume dacă interpretăm simplu elementele de operare a maşinii Turing.care efectuează algoritmul **F** drept reguli

Cu scopul de a demonta cazul **II**, ne-am putea imagina că s-ar putea aplica argumentaţia de mai sus, oricărui sistem corect $\mathbb{F}$, indiferent dacă axiomele sale sunt în număr finit sau infinit. Într-adevăr, este aşa, dar în procesul de trecere de la un sistem cu număr infinit de axiome la unul cu număr finit, trebuie introduse noi reguli de procedură care s-ar putea să nu fie, în mod evident, corecte. Astfel, atunci când, aşa cum am spus mai înainte, vom contempla axiomele şi regulile de procedură ale lui $\mathbb{F}$, aflate toate în faţa noastră, teoremele aşa zisului $\mathbb{F}$ fiind presupuse chiar acelea care sunt, în principiu, accesibile înţelegerii umane şi intuiţiei, noi nu vom putea fi siguri că regulile de procedură ale acestui $\mathbb{F}$, spre deosebire de axiomele sale, vor putea fi vreodată luate fără îndoială drept corecte, chiar dacă ar fi cu adevărat corecte. Căci, regulile de procedură nu se numără printre teoremele sistemului formal. Şi doar teoremele lui $\mathbb{F}$ pot fi acceptate, ca lucruri înţelese fără vreo urmă de îndoială.

Pentru mine nu este clar dacă această argumentaţie poate fi dusă mult mai departe în termeni strict logici. Ceea ce ar trebui să acceptăm, dacă e să credem în **II**, ar fi existenţa un anumit sistem formal $\mathbb{F}$ (care stă la baza percepţiei umane a adevărului, de exemplu a adevărului propoziţiilor π_1) care poate fi foarte bine apreciat de matematicieni, a cărui listă de axiome este (sigur) acceptabilă, dar al cărui sistem finit de reguli de procedură $\mathcal{P}$ ar conţine cel puţin o operaţie considerată ca fundamental nesigură. Toate teoremele lui $\mathbb{F}$ se vor dovedi, individual, lucruri care pot fi percepute ca adevărate într-un mod oarecum miraculos, deoarece multe dintre ele au fost obţinute prin utilizarea regulilor dubioase din $\mathcal{P}$. Fiecare dintre aceste teoreme ar putea fi *individual* percepută ca adevărată (în principiu) de către matematicieni, deşi nu există un mod *uniform* de a face acest lucru. Ne putem restrânge atenţia la acele teoreme din $\mathbb{F}$ care sunt propoziţii π_1. Prin utilizarea dubiosului $\mathcal{P}$, toată lista de propoziţii π_1 care pot fi percepute, de către un matematician, ca adevărate poate fi generată computaţional. Individual, oricare dintre propoziţiile π_1 poate fi în cele din urmă percepută ca adevărată de către intuiţia umană. Dar, în fiecare caz, acest lucru se face utilizând anumite mijloace de raţionament care sunt foarte diferite de regulile din $\mathcal{P}$, cu ajutorul cărora a fost obţinută. Alte şi alte intuiţii umane, noi şi tot mai sofisticate, trebuie introduse, astfel încât fiecare propoziţie π_1 să fie redusă la un adevăr de neatacat. Ca prin miracol, fiecare dintre aceste propoziţii π_1 se constată că sunt adevărate, dar unele dintre ele pot fi considerate adevărate doar introducând un tip

fundamental nou de raţionament, o cerinţă care intervine din nou şi din nou, la un nivel din ce în ce mai adânc. Mai mult, orice propoziţie π_1 care — indiferent prin ce mijloace — poate fi considerată indubitabil adevărată se va găsi în lista generată de $\mathcal{P}$. În final, va exista o propoziţie specială π_1, adevărată, $G(\mathbb{F})$, care poate fi construită dintr-o cunoaştere a sistemului $\mathbb{F}$, dar a cărui adevăr nu poate fi perceput ca de necontestat de către matematicieni. Cel mai bun lucru pe care ei îl pot face este să vadă dacă adevărul lui $G(\mathbb{F})$ depinde într-adevăr de corectitudinea procedurii dubioase $\mathcal{P}$, care, în mod miraculos, ar fi în stare să genereze toate propoziţiile π_1, pe care matematicienii le-ar *putea* percepe ca sigure.

Îmi imaginez că anumite persoane ar putea privi acest lucru ca acceptabil. Există multe rezultate matematice care pot fi obţinute cu ajutorul aşa-ziselor 'principii euristice', care nu furnizează o *demonstraţie* a rezultatului cerut, ci pot conduce la anticiparea faptului că rezultatul ar trebui să fie adevărat. În consecinţă, o demonstraţie s-ar putea obţine urmând mai multe linii diferite. Totuşi, mie mi se pare că astfel de principii euristice au prea puţin în comun cu presupusul nostru $\mathcal{P}$. Astfel de principii sunt lucruri care îmbunătăţesc înţelegerea noastră conştientă cu privire la „de ce anumite rezultate matematice sunt de fapt adevărate"[3] Mai târziu, când tehnicile matematice vor fi mai dezvoltate, va deveni mai clar de ce astfel de principii euristice funcţionează. Dar mai important, vor fi înţelese *circumstanţele* în care se poate avea încredere în faptul că principiul funcţionează, existând alte circumstanţe în care încrederea nu poate fi asigurată — astfel încât pot fi obţinute rezultate eronate atunci când nu există o grijă deosebită. Cu această grijă asigurată, principiul în sine devine un instrument puternic şi de încredere pentru demonstraţii matematice irefutabile. Pentru a demonstra propoziţiile π_1, decât să furnizeze un proces algoritmic care, doar în mod miraculos, este de încredere şi unde motivele pentru care algoritmul funcţionează sunt inaccesibile intuiţiei umane, principiile euristice oferă un mijloc de întărire a intuiţiilor noastre matematice şi al înţelegerii. Acest proces este foarte deosebit de algoritmul **F** (sau de sistemul

[3] Un principiu euristic de acest tip poate lua forma unei conjecturi, ca de exemplu conjectura Taiyama (care este o generalizare a 'filosofiei lui Lagland'), din care poate fi derivată, ca o consecinţă, cea mai faimoasă dintre toate propoziţiile Π_p cunoscută ca ultima teoremă a lui Fermat (vezi nota de subsol nr.9, din finalul paragrafului 3.26). Totuşi, argumentaţia prezentată de Andrew Wiles ca demonstraţie a afirmaţiei lui Fermat, nu este independentă de conjectura Taiyama - care ar fi trebuit considerată aici un sistem de reguli $\mathcal{P}$ - ci un argument pentru demonstrarea (în cazul relevant) conjecturii Taiyama însăşi!

formal $\mathcal{F}$) care este necesar pentru cazul **II**. Mai mult, nu a fost niciodată propus un principiu euristic care să poată genera cu siguranţă *toate* propoziţiile π_1 care pot fi percepute ca adevărate de către un matematician profesionist.

Evident, nimic din toate acestea nu spune că un astfel de **F** – aşa-zisa maşină de demonstrat teoreme a lui Gödel – este o imposibilitate; dar, din punctul de vedere al înţelegerii noastre matematice, existenţa sa pare foarte improbabilă. În orice caz, nu există astăzi nici cea mai slabă sugestie despre natura unui astfel de **F** plauzibil şi nici nu există vreo indicaţie despre existenţa sa. Ar putea fi doar o *conjectură*, în cel mai bun caz – şi chiar una de nedemonstrat (demonstrarea ei însemnând contrazicerea ei!). Mi se pare că ar fi o nesăbuinţă pentru orice suporter **IA** (atât pentru cei care susţin punctul de vedere $\mathcal{A}$, cât şi pentru cei care susţin $\mathcal{B}$) să lege speranţe de descoperirea unei astfel de proceduri algoritmice[4], ca cea întruchipată de **F**, a cărei existenţă este dubioasă şi care, în orice caz, dacă ar fi să existe, construirea ei efectivă ar depăşi capacitatea matematicienilor şi logicienilor din zilele noastre.

S-ar putea concepe că un astfel de **F** ar putea exista şi că ar putea fi construit cu ajutorul sau prin intermediul unei proceduri de calcul *bottom-up*, suficient de elaborate? În paragrafele 3.5-3.23, ca parte a discuţiei depre **III**, voi prezenta o argumentaţie logică riguroasă care să arate că nici o procedură *bottom-up*, care ar putea fi cunoscută, nu ar putea găsi un astfel de **F**, chiar dacă el ar exista. Ajungem astfel la concluzia că nici chiar „maşina lui Gödel pentru demonstrat teoreme" nu este o posibilitate logică serioasă, afară de cazul în care există „mecanisme incognoscibile" care stau la baza înţelegerii matematice ca întreg, lucru care nu aduce nici o consolare susţinătorilor **IA**!

Înainte de a ne ocupa de cazul **III** şi de procedura *bottom-up*, trebuie să încheiem discuţia cazului **II**; deoarece a rămas alternativa ca acţiunea algoritmică **F** – sau sistemul formal $\mathcal{F}$ – să nu fie *corect* (situaţie care nu se aplică cazului **I**). Ar putea înţelegerea matematică a oamenilor să fie echivalentă cu un algoritm cunoscut ca fiind fundamental eronat? Să examinăm această posibilitate.

---

[4] Evident, construirea unui robot matematician este foarte departe de scopurile imediate ale inteligenţei artificiale şi, în consecinţă, găsirea unui astfel de **F** poate fi privită ca prematură sau inutilă. Totuşi, acest lucru ar însemna să scăpăm esenţa acestei discuţii. Acele puncte de vedere care consideră inteligenţa umană explicabilă în termenii unui proces algoritmic, cer, implicit, prezenţa potenţială a unui astfel de **F** – cognoscibil sau incognoscibil – deoarece chiar aplicarea inteligenţei este cea care ne-a condus la concluziile noastre. Nu există nimic special aici privind abilităţile matematice, în această privinţă vezi paragrafele 1.18 şi 1.19.

3.4. Utilizează oare, fără discernământ, matematicienii un algoritm incorect ?

Poate că *există* un sistem formal $\mathcal{F}$ care, fără a fi *corect*, stă la baza înţelegerii noastre matematice. Cum putem fi siguri că percepţiile noastre matematice despre ce este cu siguranţă adevărat nu ar putea într-o bună zi să ne ducă pe un drum fundamental greşit ? Probabil, aşa ceva s-a şi întâmplat deja. Aceasta nu este chiar situaţia pe care am considera-o în legătură cu **I**, unde am exclus posibilitatea de a cunoaşte faptul că un sistem $\mathcal{F}$ ar putea juca acest rol. Aici convenim că acest *rol* al lui $\mathcal{F}$ nu poate fi cunoscut şi astfel trebuie să acceptăm ipoteza că sistemul $\mathcal{F}$ nu ar fi corect. Dar este oare plauzibil să acceptăm că inatacabilele noastre convingeri matematice se bazează pe un sistem care nu ar fi corect atât de incorect încât '1=2' ar face parte din ceea ce el implică? Cu siguranţă, dacă nu putem avea încredere în raţionamentul matematic, atunci nici un fel de raţionament despre modul în care funcţionează lumea nu mai este de încredere. Pentru că raţionamentul matematic formează o parte esenţială a întregii noastre înţelegeri ştiinţifice.

Totuşi, unii ar putea susţine că nu este de neconceput ca raţionamentul nostru matematic general acceptat (sau ceea ce ar putea, în viitor, deveni de neatacat) ar putea conţine unele contradicţii interne, ascunse. Astfel de persoane s-ar putea referi la paradoxul celebru (despre „mulţimea mulţimilor care nu se conţin ca element") pe care Bertrand Russell l-a indicat într-o scrisoare către Gottlab Frege, în 1902, tocmai pe când Frege se pregătea să-şi publice opera vieţii sale despre fundamentele matematicii (vezi şi răspunsul din paragraful 2.7 la **Q9** precum şi **MN** p.115). Frege adăuga în apendixul lucrării (Frege, traducerea engleză, 1964):

> „Greu s-ar putea întâmpla unui om de ştiinţă ceva mai rău decât ca fundamentele edificiului pe care l-a creat să fie zguduite după ce munca sa este încheiată. Eu m-am aflat într-o astfel de situaţie după primirea unei scrisori din partea domnului Bertrand Russell..."

Bineînţeles, am putea spune că, pur şi simplu, Frege a făcut o greşală. Se

admite, în general, că matematicienii pot, din când în când, face câte o greşeală — uneori chiar serioasă. Mai mult, eroarea lui Frege era *corectabilă*, după cum rezultă şi din recunoaşterea lui Frege. N-am arătat, chiar eu, aici, că astfel de greşeli corectabile nu le vom lua în considerare? (vezi, paragraful 2.10, răspunsul la **Q13**). Aşa cum spuneam acolo, noi suntem cu adevărat preocupaţi de chestiunea de principiu şi nu de scăpările unui anumit matematician. Este oare sigur că erorile care pot fi depistate şi demonstrate ca erori sunt lucruri care ar trebui să nu ne preocupe aici? Trebuie spus că situaţia de acum este diferită de cea de la prezentarea răspunsului la **Q13**, deoarece acum suntem preocupaţi de un sistem formal F despre care *nu ştim* că ar sta la baza înţelegerii matematice. Ca mai înainte, nu ne preocupăm de greşelile individuale — sau 'scăpările' pe care unii matematicieni le-ar putea face într-un raţionament bazat pe o schemă consistentă. Acum ne preocupăm de situaţia în care schema însăşi ar putea să fie subiectul unei contradicţii. Acest lucru s-a întâmplat în cazul lui Frege. Dacă paradoxul lui Russell, sau un alt paradox de aceiaşi natură, nu ar fi fost descoperit şi prezentat lui Frege, atunci s-ar fi putut întâmpla ca acesta să nu observe că schema sa conţinea o eroare fundamentală. Nu era vorba ca Russell să fi indicat o scăpare tehnică în raţionamentul lui Frege, pe care Frege să o fi acceptat, ca eroare, în concordanţă cu propria sa metodă de raţionament; se întâmplase că tocmai metoda sa se dovedea că are în sine o contradicţie. Chiar această *contradicţie* îi indica lui Frege că exista o eroare şi, ceea ce până atunci părea de neatacat, se dovedea că are un defect fundamental. Dar defectul a fost perceput deoarece contradicţia însăşi a fost pusă în lumină. Fără ca această contradicţie să devină evidentă, metoda de raţionament ar fi fost în continuare considerată infailibilă şi probabil urmată de matematicieni încă o bună bucată de timp.

În fond, trebuie să spun că, în acest caz, este greu de crezut că matematicienii şi-ar fi permis încă multă vreme libertatea de a raţiona (cu mulţimi infinite) aşa cum permitea schema lui Frege. Dar acest lucru se datorează faptului că paradoxuri de tipul paradoxului lui Russell sunt uşor de descoperit. Ne-am putea imagina însă nişte paradoxuri mult mai subtile, aflate implicit chiar în ceea ce credem că este o procedură matematică de neatacat pe care o folosim astăzi — un paradox care ar putea ieşi la lumină peste câteva secole. Doar când un astfel de paradox s-ar manifesta, ar apărea necesitatea de a modifica regulile actuale. Un argument ar fi că intuiţiile noastre matematice nu sunt guvernate numai de

considerații eterne, ci și că aceste intuiții se modifică, influențate fiind de ceea ce pare că funcționează bine *deocamdată* și cu care, de fapt, „ne-am descurcat foarte bine" Văzând lucrurile în acest mod, s-ar putea accepta că ar putea exista un algoritm sau un sistem formal care să se afle în prezent la baza înțelegerii matematice, dar care să nu fie un lucru fix, fiind tot timpul supus la schimbări, pe măsură ce apar noi informații. Voi reveni ceva mai târziu (în paragrafele 3.9-3.11, vezi și 1.5) la acest algoritm schimbător. Vom vedea că astfel de lucruri sunt tot algoritmi, dar într-un veșmânt diferit.

Evident, ar fi naiv din partea mea să nu accept că, în modul în care operează matematicienii în practică există adeseori elemente de „încredere într-o procedură, dacă pare că ea a fost utilizabilă până acum" Dar astfel de lucruri apar atunci când se bâjbâie în căutarea unei înțelegeri care nu a fost încă obținută și nu se aplică lucrurilor deja stabilite indubitabil. Cred că nici Frege însuși nu era complet dogmatic în credința lui că schema sa era infailibilă, chiar fără ca paradoxul lui Russell să-i fi fost adus la cunoștință. O astfel de schemă generală de raționament trebuie, în cel mai bun caz, luată ca o propunere. Este nevoie de o lungă perioadă de meditație asupra unui rezultat înainte ca el să poată fi luat drept ceva ajuns la nivelul în care devine de neatacat. Pentru scheme de generalitatea schemei lui Frege, în mod normal suntem conduși să formulăm propoziții de forma „presupunând că schema lui Frege este corectă atunci, așa și așa..." și nu pur și simplu să susținem proprietatea 'așa și așa' fără altă precizare (vezi comentariile **Q11, Q12.**)

Probabil matematicienii au devenit mai prudenți cu ceea ce sunt dispuși să considere 'fără îndoială adevărat' — după o perioadă de excesivă îndrăzneală (din care rezultatele lui Frege fac cu siguranță parte), la sfârșitul secolului al nouăsprezecelea. Îndată ce au apărut paradoxurile, ca cel al lui Russell, importanța acestei prudențe s-a manifestat din plin. Îndrăzneala a apărut atunci când puternica teorie a numerelor infinite și a mulțimilor infinite a lui Cantor, dezvoltată la sfârșitul secolului XIX, s-a impus (totuși, trebuie spus că, însuși Cantor era preocupat de probleme de tipul paradoxului lui Russell — cu mult înainte ca Russell să îl fi găsit [4] — și a încercat să formuleze un punct de vedere foarte sofisticat care să ia în considerare astfel de lucruri). Pentru tot ceea ce discutăm noi aici, este necesară o precauție deosebită. Sunt fericit că discutăm numai despre lucruri a căror adevăr este de nezdruncinat și că probleme legate de mulțimile

infinite nu vor fi incluse. Esențial este că, în orice caz, argumentul lui Gödel produce afirmații care rămân în interiorul lucrurilor incontestabil de neatacat (vezi și cometariul la **Q13**). Raționamentul lui Gödel (și Turing) nu implică propriu zis existența problematică a vreunei mulțimi infinite. Chestiunile controversate în legătură cu tipul de raționament liber utilizat de Cantor, Frege și Russell nu ne preocupă atâta vreme cât rămâne 'dubios' în opoziție cu ceea ce este 'de neatacat'. Acestea fiind admise, eu nu văd cum s-ar putea crede că matematicienii ar folosi în *realitate* un sistem formal $\mathbb{F}$ care nu este *corect* ca bază pentru înțelegerea și convingerile lor matematice. Sper că și cititorul este de acord cu mine că, dacă astfel de considerații sunt posibile, ele nu sunt *plauzibile*.

În sfârșit, în legătură cu presupusul nostru sistem incorect $\mathbb{F}$ trebuie să reamintesc celelalte aspecte ale lipsei de acurateță proprii omului, despre care am discutat mai sus, în legătură cu **Q12** și **Q13**. Trebuie mai întâi să accentuez din nou că nu inspirația sau criteriile euristice, care ar putea ghida matematicienii în căutarea unor noi descoperiri, ne preocupă acum, ci înțelegerea și intuiția care furnizează baza convingerilor lor de nezdruncinat legate de adevărul matematic. Aceste convingeri pot să se formeze pur și simplu urmărind raționamentul cuiva, fără să necesite vreun element de descoperire matematică.

Atunci când cineva urmează drumul spre descoperiri originale este important să-i fie permise speculațiile libere, neconstrânse de la început de precizie și exactitate (și aceasta este impresia mea despre remarca lui Turing, pe care am citat-o mai sus, în paragraful 3.1). Dar atunci când se ajunge în punctul în care trebuie decis dacă un raționament, în sprijinul unei propoziții matematice propuse ca fiind indubitabil adevărată, trebuie acceptat sau respins, este necesar să ne preocupăm de înțelegerea și intuiția matematică — adeseori ajutată de calcule laborioase — și să ne convingem că nu conțin erori.

Acest lucru nu exclude posibilitatea ca frecvent matematicienii să se înșele în privința faptului că au aplicat corect lucrurile pe care le-au înțeles. Matematicienii fac greșeli în raționamentele lor atât în ceea ce înțeleg, cât și în calculele pe care le folosesc. Dar tendința de a face astfel de erori nu le *ridică* puterea de înțelegere (deși, pot să-mi imaginez că un fulger ocazional de înțelegere pot avea și acești oameni norocoși). Mai important decât orice, aceste erori sunt *corectabile*; atunci când astfel de erori sunt semnalate, de alți matematicieni, sau chiar de autorul lor, cândva mai târziu, erorile sunt *recunoscute* ca erori. Nu este

ca atunci când există erori interne în sistemul formal $\mathbb{F}$ care controlează înţelegerea matematică, deoarece, în acest caz, sistemul nu este capabil să-şi recunoască propriile erori. (Posibilitatea ca un sistem să fie autocorector, putând să se schimbe de câte ori descoperă vreo greşeală, va fi inclusă în discuţia pe care o vom purta în paragraful 3.14. Ca urmare, vom vedea că acest tip de propunere nu ne ajută de fapt; vezi şi 3.26.)

O situaţie diferită apare atunci când propoziţiile matematice sunt incorect formulate; propunând un rezultat, matematicianul ar putea să *înţeleagă* altceva decât ceea ce literal a spus. Din nou suntem în situaţia în care eroarea este corectabilă şi nu în faţa unei erori interne care s-ar datora unui sistem $\mathbb{F}$ care nu este *corect* şi totuşi stă la baza intuiţiilor umane. (Să ne amintim povestea despre Feynman, la care ne-am referit în legătură cu **Q13**: „Nu asculta ce spun; încearcă să pricepi ce am vrut să spun!") Noi suntem preocupaţi de ceea ce poate fi stabilit de un matematician, *în principiu* şi de aceea erorile de tipul celor pe care le-am considerat anterior — adică erorile corectabile — nu sunt relevante.

Mai importantă pentru întreaga noastră discuţie este ideea centrală din raţionamentul Gödel Turing care face parte din ceea ce matematicienii pot înţelege şi ea este cea care ne obligă să respingem **I** şi să privim cu suspiciune **II**, ca extrem de puţin probabil.

Aşa cum am spus mai sus, în discuţia relativ la **Q13**, evident că *ideea* raţionamentului Gödel-Turing ar trebui să facă parte din ceea ce un matematician poate, în principiu să înţeleagă, chiar dacă anumite enunţuri particulare $G(\mathbb{F})$, pe care matematicianul s-ar întâmpla să le stabilească, ar putea fi eronate — deoarece ele sunt *corectabile*.

Există şi alte aspecte care trebuiesc analizate în legătură cu posibilitatea ca un algoritm care nu este 'corect' să stea la baza înţelegerii matematice. Acestea sunt în legătură cu procedurile *bottom-up*; de acest tip sunt algoritmii auto-corectabili, algoritmii de învăţat (incluzând reţelele neuronale), algoritmi cu ingrediente aleatoare şi algoritmi ale căror acţiuni depind de mediul exterior în care dispozitivul este plasat. Unele dintre aceste probleme au fost atinse mai înainte (vezi comentariul la **Q2**) şi vor fi dezvoltate pe larg în discuţia care urmează, asupra cazului **III**.

3.5. Poate un algoritm să fie incognoscibil?

În conformitate cu **III,** înţelegerea matematică ar putea fi rezultatul unui anumit algoritm care nu poate fi cunoscut. Ce înseamnă de fapt pentru un algoritm să fie 'incognoscibil'? În secţiunile precedente din capitolul de faţă, ne-am preocupat de chestiuni *de principiu.* Astfel, o afirmaţie de felul celei care spune că adevărul de necontestat al unei propoziţii π_1 este accesibil înţelegerii matematice a omului, înseamnă că π_1 este o propoziţie *în principiu* accesibilă şi nu că orice matematician trebuie, în mod necesar, să găsească pentru ea o demonstraţie. Dar pentru a vedea ce înseamnă că un algoritm nu poate fi cunoscut, este necesară explicarea modului în care folosim termenul 'necognoscibil'. Voi folosi acest termen relativ la un algoritm a cărui specificare este dincolo de ceea ce se poate atinge *în practică.*

Atunci când ne-am preocupat de derivare în interiorul unui anumit sistem formal cognoscibil, sau de ceea ce se poate obţine utilizând un algoritm cunoscut, aceste lucruri priveau ceea ce poate şi ce nu poate fi obţinut în principiu. Dacă o anumită propoziţie particulară poate fi derivată dintr-un sistem formal sau un algoritm, problema este cu necesitate considerată 'în principiu' Putem compara această situaţie cu aceea a adevărului propoziţiilor π_1.

O propoziţie π_1 este, în cele din urmă, considerată *adevărată* doar dacă reprezintă acţiunea unei maşini Turing care în principiu nu se termină (acest lucru este în concordanţă cu discuţia de la **Q8**). La fel, aserţiunea că o anumită propoziţie este sau nu derivabilă într-un anumit sistem formal trebuie considerată 'în principiu', ca o aserţiune care este ea însăşi afirmaţia că o anumită propoziţie particulară π_1 este respectiv adevărată sau falsă (vezi şi sfârşitul discuţiei de la **Q10**). În consecinţă, când ne ocupăm de derivabilitatea dintr-o mulţime de reguli formale, 'cognoscibilitatea' va fi adeseori considerată 'în principiu', în sensul de mai sus. Pe de altă parte, când vrem să ştim dacă regulile în cauză pot fi *cunoscute,* această problemă se pune în sens *practic.* Orice sistem formal (maşină Turing sau propoziţie π_1) poate fi specificat în principiu, aşa încât problema 'cognoscibilităţii' este aici legată de ceea ce poate sau nu poate fi descris. Orice algoritm, de orice fel, este în principiu cognoscibil — în sensul că acţiunea maşinii Turing care efectuează algoritmul este 'cunoscută', îndată ce numărul natural care codifică acţiunea este cunoscut (în termenii numerelor care specifică maşina

Turing, aşa cum am arătat în **MN**).

Nu există nici o sugestie că un număr natural ar putea fi necognoscibil în principiu. Numerele naturale (şi astfel acţiunile algoritmilor) pot fi scrise 0, 1, 2, 3, 4, 5, 6, şi astfel orice număr natural poate fi atins prin enumerare — în principiu — indiferent cât de mare ar fi acest număr! În practică, însă, există numere atât de mari încât nu avem nici o speranţă să le atingem prin enumerare. De exemplu, numărul maşinii Turing descrisă în **MN**, p. 66, este mult prea mare ca să poată fi atins practic, la o astfel de enumerare. Chiar dacă numerele naturale ar putea fi produse unul după altul într-un interval tot atât de mic de timp, cât poate fi teoretic definit (scala de timp a lui Planck este de 0.5×10^{-43} secunde — vezi paragraful 6.11), totuşi nici un număr având o reprezentare binară cu mai mult de 203 cifre nu ar putea fi atins prin enumerare, în tot timpul care s-a scurs de la începutul universului, începând cu *big-bang*-ul. Numărul specificând maşina Turing discutată mai înainte, are cam de 20 de ori mai multe cifre — dar nici despre acesta nu se poate spune că ar fi 'necognoscibil' în practică, dat fiind că este numărul prezentat explicit în **MN**.

Pentru ca un număr natural sau o acţiune a unei maşini Turing să fie 'necognoscibilă' în practică, trebuie să ne imaginăm că şi specificarea numărului ar fi atât de complicată încât s-ar afla dincolo de capacitatea omenească. S-ar putea spune că este un ordin de mărime foarte mare. Ţinând seamă de finitudinea fiinţelor umane, trebuie să existe anumite limite care pun anumite numere deasupra capacităţii umane (vezi şi discuţia de la răspunsul **Q8**).

Ne putem imagina, în concordanţă cu **III**, cât de mare ar fi complicaţia tuturor detaliilor specificării unui algoritm **F** despre care s-ar presupune că stă la baza înţelegerii umane, care să-l pună deasupra posibilităţilor de cunoaştere — în sensul posibilităţii de a fi descris — prin contrast cu cognoscibilitatea algoritmului pe care se presupune că îl utilizăm de fapt. Această problemă a posibilităţii specificării algoritmului desparte **III** de **II**. Astfel, în analiza lui **III** trebuie să luăm în considerare situaţia în care specificarea numărului ar fi deasupra capacităţii umane, fără a mai discuta despre proprietăţile sale ca număr care determină acţiunea algoritmică care stă la baza înţelegerii matematice a oamenilor.

Trebuie totuşi să fie clar că doar mărimea nu poate fi un factor de limitare. Sunt uşor de specificat numere atât de mari încât le *întrec* pe acelea care ar putea fi necesare pentru a specifica o acţiune algoritmică relevantă pentru comportarea

unui organism în universul observabil (de exemplu, cât de uşor este specificat numărul 2^{65536}, care apare în răspunsul la **Q8** şi care depăşeşte cu mult numărul tuturor stărilor diferite ale universului pentru toată materia care se află în universul nostru observabil [5]). Ar fi vorba despre specificarea *precisă* a numărului cerut şi nu despre mărimea lui, atunci când spunem că depăşeşte capacitatea umană.

Să presupunem că, în concordanţă cu **III**, specificarea unui astfel de **F** ar fi peste capacităţile umane. Ce ne va spune acest lucru despre posibilitatea unei stategii **IA** de succes deplin (conform cu **IA** 'tare' sau **IA** 'slabă' — respectiv cu punctele de vedere $\mathcal{A}$ sau $\mathcal{B}$) ?

Se va anticipa de către susţinătorii sistemelor **IA** (cu certitudine cei ce susţin $\mathcal{A}$ dar şi cei ce susţin $\mathcal{B}$) că roboţii care ar putea eventual să fie realizaţi ca rezultat al acestei strategii vor trebui să atingă şi, probabil, să depăşească, posibilităţile matematice ale omului. Ca o consecinţă, s-ar putea, dacă acceptăm **III**, ca un astfel de algoritm **F**, care omeneşte este de nespecificat, să facă parte din sistemul de comandă al unui astfel de robot matematic. Acest lucru pare să implice faptul că o strategie **IA** cu un astfel de scop este de neatins. Pentru că dacă are nevoie, pentru a-şi atinge scopul, de un algoritm **F** care nu poate fi specificat, atunci nu ar exista nici o speranţă pentru fiinţele umane să-l poată pune în funcţiune vreodată.

Dar nu acesta este tabloul prezentat de cei mai ambiţioşi suporteri **IA**. Ei ar putea să susţină că algoritmul necesar, **F**, poate să nu fie clar de la început şi că el s-ar construi în etape, în care robotul însuşi şi-ar îmbunătăţi performanţele prin învăţare din experienţă. Mai mult, roboţii cei mai avansaţi nu vor fi creaţia directă a oamenilor ci, mai degrabă, vor fi creaţi de alţi roboţi [6], probabil unii mai primitivi dacât ceea ce se cere robotului matematic; şi s-ar putea imagina o evoluţie *darwinistă* a operaţiei, folosind la îmbunătăţirea capacităţilor, de la o generaţie la alta. Tot astfel, s-ar putea spune că, printr-un proces general de acelaşi tip, noi înşine am fost capabili să câştigăm, ca ingredienţi ai 'calculatorului nostru neuronal', vreun algoritm **F** care nu poate fi cunoscut şi care comandă înţelegerea noastră matematică.

În secţiunile următoare, voi arăta că procese de această natură nu ne scapă de următoarea problemă: dacă procedurile, pe care o strategie **IA** le-ar indica, ar fi algoritmice şi cognoscibile, atunci orice sistem **F**, astfel obţinut, ar trebui să fie şi el cognoscibil. În acest fel, cazul **III** se va reduce sau la **I** sau la **II**, cazuri care

au fost desfiinţate în paragrafele 3.2-3.4, ca fiind efectiv imposibile (cazul I), sau extrem de neplauzibile (cazul II). De fapt, suntem conduşi la cazul I dacă presupunem că o astfel de procedură algoritmică poate fi cunoscută. În consecinţă, cazul III (şi prin implicaţie şi cazul II) ar trebui el însuşi tratat ca efectiv de ne-apărat.

Orice cititor care crede cu tărie în posibilitatea ca III să furnizeze o metodă verosimilă pentru un model computaţional al minţii, va trebui să parcurgă aceste argumente cu atenţie şi să le urmărească cu cea mai mare grijă.

Concluzia va fi că, dacă se consideră că III furnizează baza pentru propria noastră înţelegere matematică, atunci singura metodă plauzibilă ca propriul nostru F să fi apărut, ar fi fost cu ajutorul intervenţiei divine — în fond posibilitatea $\mathcal{AD}$ pe care am menţionat-o la sfârşitul paragrafului 1.3 — şi acest lucru nu ar fi o consolare pentru cei preocupaţi de scopuri IA mai ambiţioase, pe termen lung!

3.6. Selecţie naturală sau acţiunea lui Dumnezeu?

Probabil că ar trebui să considerăm cu seriozitate posibilitatea ca inteligenţa noastră să necesite cu adevărat o anumită acţiune a lui Dumnezeu — şi acest lucru nu poate fi explicat în termenii ştiinţei, cea care a avut un succes atât de mare în descrierea lumii neînsufleţite. Evident, vom continua să rămânem cu mintea deschisă; dar eu trebuie să declar că în discuţia care urmează, voi susţine un punct de vedere ştiinţific. Voi analiza atât posibilitatea ca înţelegerea noastră matematică să rezulte dintr-un algoritm de nepătrus — cât şi întrebarea: cum apare un astfel de algoritm — în întregime în termeni ştiinţifici. S-ar putea ca unii cititori, să fie înclinaţi să creadă că un astfel de algoritm ar fi putut pur şi simplu să fi fost implantat în creierul nostru printr-un act divin. Această sugestie nu o pot respinge decisiv; dar dacă cineva alege să abandoneze metoda ştiinţifică la un moment dat, pentru mine nu este clar de ce i s-ar fi părut că acesta ar fi momentul potrivit! Dacă metoda ştiinţifică este să fie abandonată, nu ar fi mai potrivit ca sufletul să fie eliberat de orice acţiune algoritmică şi să nu i se obtureze presupusul liber arbitru în complicaţiile şi insondabilitatea unui algoritm despre care se presupune că îi comandă toate acţiunile?

Într-adevăr, pare mai simplu să adoptăm poziţia pe care Gödel însuşi a

susţinut-o şi anume că activitatea mentală ar trebui să se plaseze undeva deasupra funcţionării creierului fizic — ceea ce ar fi pe linia punctului de vedere $\mathcal{D}$. Pe de altă parte, îmi închipiu că în zilele noastre nici cei care susţin că într-un anumit sens intelectul nostru (activitatea noastră mentală) este un dar divin, nu pot să nu aibă tendinţa de a explica, cu metode ştiinţifice, comportamentul uman. Fără îndoială, aceste poziţii pot fi discutate, dar scopul meu, în acest moment, nu este de a combate punctul de vedere $\mathcal{D}$. Sper că acei cititori care susţin punctul de vedere $\mathcal{D}$, într-o formă sau alta, mă vor tolera şi vor încerca, împreună cu mine, să vadă ce fel de argumente ştiinţifice am putea aduce în sprijinul nostru.

Care sunt implicaţiile presupunerii că ajungerea la concluzii matematice se face ca rezultat al acţinunii unui algoritm cu necesitate insondabil? În mare, tabloul ar trebui să fie acela al unor excepţional de complicate procese algoritmice care pentru a simula înţelegerea matematică pură, ar fi avut nevoie de câteva sute de mii de ani, (cel puţin) de selecţie naturală şi de câteva mii de ani de educaţie tradiţională şi de comunicare cu mediul fizic ambiant.

Aspectele ereditare ale acestor proceduri se vor fi construit probabil pornind de la ingredienţi algoritmici mai simpli (şi mai primitivi), ca rezultat al acelor presiuni de selecţie care au avut ca efect şi celelate componente ale superbei maşinării care este corpul nostru şi mintea noastră. Potenţialii algoritmi matematici, ereditari (adică aspectele moştenite ale gândirii noastre matematice — algoritmice, sau de altă natură) ar trebui să fie codificaţi în ADN-ul nostru, ca particularităţi ale caracteristicilor secvenţelor de nucleoizi şi ele ar fi trebuit să apară ca rezultat al unei proces prin care perfecţionarea gradată sau intermitentă se face ca răspuns la presiunea selecţiei naturale. În plus, ar trebui să existe influenţe externe de diverse tipuri, ca educaţia matematică, experienţa câştigată din mediul nostru fizic şi alţi factori aleatori de influenţare. Trebuie să vedem dacă un astfel de tablou poate fi considerat plauzibil.

3.7. Un algoritm sau mai mulţi algoritmi?

Problema importantă pe care trebuie să ne-o punem este dacă ar putea exista mai mulţi algoritmi, diferiţi şi probabil neechivalenţi, care ar fi responsabili de faptul că diferiţi indivizi posedă diferite tipuri de înţelegere matematică. Într-

adevăr, un lucru este clar dintru început şi anume că indivizi diferiţi chiar dintre matematicienii profesionişti percep matematica în moduri diferite unii de alţii. Pentru unii, imaginea vizuală este foarte importantă, în vreme ce pentru alţii, trebuie să fie precizată structura logică, argumente conceptuale subtile, sau raţionamente analitice detaliate, sau simple manipulări ale calculelor algebrice. În legătură cu acest lucru este important să amintim că gândirea geometrică şi analitică se crede că au loc în lobul drept, respectiv stâng, al creierului [7]. Deşi acelaşi adevăr matematic poate adesea să fie perceput în oricare dintre aceste moduri.

Din punct de vedere algoritmic, ar putea să pară că există o neechivalenţă profundă între diferiţii algoritmi matematici pe care diverse persoane îi posedă. Dar, în ciuda imaginilor foarte diferite pe care şi le pot forma diferiţi matematicieni (sau alte persoane) cu scopul de a înţelege sau a comunica idei matematice, un fapt surprinzător în legătură cu percepţiile matematice este că atunci când matematicienii prezintă ceea ce ei cred că este indubitabil adevărat, nu apar dezacorduri, cu excepţia cazurilor în care dezacordurile se datorează unor eventuale erori de raţionament (corectabile), pe care autorul le recunoaşte, sau eventual în legătură cu un număr foarte mic de probleme fundamentale ca cele discutate în **Q11**, în particular $\mathcal{G}^{**}$ Pentru comoditate voi ignora în discuţia care urmează, acest ultim aspect. Deşi are o anumită importanţă, acest aspect nu afectează în mod substanţial concluziile noastre (având un număr foarte mic de puncte de vedere diferite, acest lucru nu se deosebeşte esenţial de cazul în care ar exista un unic punct de vedere).

Percepţia adevărului matematic se poate atinge în multe moduri. Nu există nici o îndoială că oricare ar fi activitatea fizică ce are loc atunci când o persoană percepe un adevăr matematic, această activitate fizică diferă substanţial de la un individ la altul, chiar dacă ei percep acelaşi adevăr matematic.

Aşadar, dacă matematicienii ar folosi algoritmi matematici pentru a-şi forma judecăţile despre adevărul matematic indubitabil, chiar aceşti algoritmi ar trebui să difere, în structura lor, de la un individ la altul. Deci într-un anumit sens, algoritmii ar trebui să fie echivalenţi unul cu altul.

Acest lucru nu este de neînţeles, aşa cum ar putea să pară la prima vedere, atâta vreme cât poate fi considerat matematic că este *posibil*. Maşini Turing, foarte diferite au aceleaşi date de ieşire (de exemplu, să considerăm maşina Turing

construită în următorul fel: atunci când acţionează asupra unui număr natural n scoate la ieşire un 0 dacă n poate fi reprezentat ca sumă a patru pătrate şi 1 în caz contrar. Datele de la ieşirea acestei maşini sunt identice cu cele ale unei alte maşini, construită astfel încât, să scoată la ieşire valoarea 0, *ori de câte ori* acţionează asupra unui număr n — acest lucru se întâmplă deoarece orice număr natural n este suma a patru pătrate; vezi paragraful 2.3).

Doi algoritmi nu trebuie să fie similari în ceea ce priveşte construcţia lor internă şi, totuşi, pot fi identici în ceea ce priveşte eventualele lor efecte externe. Totuşi, într-un anumit sens, este încă mai uimitor cum ar fi putut să apară aceşti algoritmi insondabili destinaţi stabilirii adevărului matematic, căci acum avem nevoie de mai mulţi algoritmi, diferiţi unul de altul în privinţa detaliilor construcţiei, dar totuşi echivalenţi din punctul de vedere al valorilor la ieşire.

3.8. Selecţia naturală a matematicienilor

Ce se poate spune despre rolul selecţiei naturale ? S-ar putea să existe un algoritm (sau mai mulţi) care, controlând înţelegerea noastră matematică, să fie necognoscibili (aşa cum ar rezulta din **III** — sau care să joace un rol care nu poate fi cunoscut (cum susţine **II**)? Permite-ţi-mi să reamintesc o observaţie care a fost făcută la începutul paragrafului 3.1. Matematicienii nu cred că atunci când ajung la ceea ce consideră a fi concluzii matematice de nezdruncinat, ar urma pentru aceasta un set de reguli necunoscute, reguli care să intervină într-un asemenea mod încât să fie în principiu, matematic, insondabile.Ei cred că aceste concluzii conduc, dimpotrivă, în ciuda demonstraţiilor adesea lungi şi chinuitoare, la adevăruri clare şi de necontestat care pot fi apreciate, în principiu, de oricine.

Într-adevăr, la nivelul simţului comun sau într-o descriere logică, ceea ce ei cred că fac, este chiar ceea ce fac. Acest lucru nu poate fi pus la îndoială; este un aspect pe care nu pot să-l subliniez destul. Dacă ar fi să ne menţinem pe poziţia care susţine că matematicienii urmăresc un set necunoscut şi insondabil de reguli computaţionale, aşa cum rezultă din **II** şi **III**, atunci acest lucru ar trebui să fie ceva ce ei fac — simultan cu ceea ce cred că fac, dar la nivele diferite de descriere. Într-un anumit fel,urmărirea acelor reguli ar trebui să aibă, în practică cel puţin, acelaşi *rezultat*, ca şi rezultatul înţelegerii matematice şi al intuiţiei. Ceea

ce trebuie să încercăm să credem, dacă aderăm la unul dintre punctele de vedere $\mathcal{A}$ sau $\mathcal{B}$, ar fi că lucrul acesta este plauzibil.

Să ţinem minte că aceşti algoritmi trebuie să realizeze scopul propus. Algoritmii trebuie să-i înzestreze pe posesorii lor cu capacitatea ca, cel puţin în principiu, să poată urmări raţionamente matematice asupra unor entităţi abstracte foarte îndepărtate de experienţa directă şi care, în cea mai mare parte, nu oferă persoanelor care îi posedă nici un avantaj practic. Oricine aruncă o privire într-o revistă de cercetare matematică modernă, va înţelege cât de departe de orice lucru practic sunt preocupările matematicienilor. Detaliile demonstraţiilor prezentate în astfel de lucrări nu sunt complet inteligibile decât pentru o mică minoritate de cercetători; dar raţionamentele sunt obţinute prin paşi mici, fiecare în parte poate fi înţeles, în principiu, de către orice persoană, chiar atunci când privesc proprietăţi abstracte asupra unor mulţimi infinite complicate. Trebuie să presupunem că natura unei anumite secvenţe ADN procură un algoritm potrivit (sau, mai probabil, un număr mare de algoritmi matematic echivalenţi) pentru a da oamenilor capacitatea de a urmări astfel de raţionamente. Dacă credem aşa ceva, atunci trebuie să ne întrebăm cu toată seriozitatea cum de un astfel de algoritm (sau set de algoritmi) a putut să apară prin selecţie naturală.

Este evident că astăzi nu este nici un avantaj, din punctul de vedere al selecţiei, să faci matematică. (Ba, aş spune chiar că este un dezavantaj. Matematicienii cu înclinaţii puriste au tendinţa să sfârşească în poziţii academice prost plătite — sau să rămână chiar fără slujbă — ca rezultat al pasiunii lor ciudate!) Încă mai evident, nu exista nici un avantaj în selecţie pentru strămoşii noştri îndepărtaţi să poată raţiona cu mulţimi infinite, foarte abstracte şi cu mulţimi infinite de mulţimi infinite, etc. Acei strămoşi erau preocupaţi de problemele practice de zi cu zi; de lucrări de tipul construcţiei adăpostului, confecţionării hainelor sau proiectării capcanelor pentru mamuţi — sau, mai târziu, cele de domesticire a animalelor şi de cultivare a grâului (ca în figura 3.1.).

Ar fi foarte rezonabil să presupunem că avantajele la selecţie pe care strămoşii noştri le considerau de dorit, erau acele calităţi care se potriveau acestor lucruri şi doar o trăsătură *incidentală* s-a dovedit, mult mai târziu, exact aceea de care era nevoie pentru ducerea la bun sfârşit a raţionamentelor matematice. Cam asta este, mai mult sau mai puţin, ceea ce cred şi eu. Într-o astfel de viziune, posibilitatea de *a înţelege* a fost calitatea generală pe care Omul a câştigat-o într-un

fel, sau a dezvoltat-o sub presiunea selecţiei naturale. Această capacitatea de înţelegere a lucrurilor nu era specifică şi a fost utilizată în multe feluri de Om în avantajul său. Construcţia adăposturilor, a capcanelor pentru mamuţi, de exemplu, erau ocazii în care capacitatea de înţelegere a Omului era inestimabilă. După părerea mea, capacitatea de a înţelege ar fi, fără îndoială, o calitate unică pentru Homo Sapiens. Ea trebuia să fie prezentă la multe alte animale cu care Omul a intrat în competiţie, dar într-o măsură mai mică, aşa încât Omul, în virtutea *măririi* capacităţii sale de înţelegere a obţinut la selecţie un avantaj considerabil.

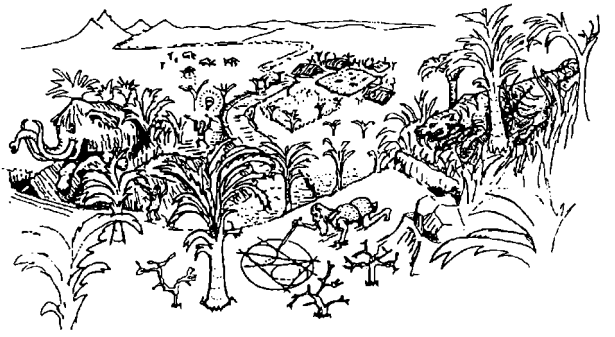


Fig. 3.1. E greu de crezut că pentru strămoşii noştri îndepărtaţi, capacitatea specială de a face matematică putea fi un avantaj în selecţia naturală

Dificultatea cu un astfel de punct de vedere apare numai dacă încercăm să ne imaginăm că facultatea înnăscută de a înţelege ar fi 'algoritmică' Deoarece, aşa cum am văzut, pe baza argumentelor pe care le-am adus mai sus, orice facultate de înţelegere care este suficient de puternică pentru a permite posesorului său să înţeleagă argumentele matematice şi, în particular, demonstraţia lui Gödel, în forma pe care am dat-o aici, trebuie, dacă este algoritmică, să fie atât de complicată încât ea (sau rolul ei) să nu poată fi cunoscută nici chiar de posesorul acestei facultăţi.

Presupusul algoritm obţinut prin selecţie naturală ar fi trebuit să fie suficient de puternic încât în vremea strămoşilor noştri îndepărtaţi să fi conţinut deja, potenţial, regulile oricărui sistem formal care este considerat astăzi, de matematicieni, indubitabil consistent (sau inatacabil corect, în privinţa propoziţiilor π_1, conform cu răspunsul **Q10** din parafraful 2.20). Acesta ar fi inclus, aproape

sigur, regulile sistemului formal Zermelo-Frenkel, $\mathbb{ZF}$, sau poate ale sistemului extins $\mathbb{ZFC}$ (care este $\mathbb{ZF}$, la care se adaugă axioma alegerii) — sistem pe care mulți matematicieni îl consideră că oferă toate metodele de raționament necesare pentru matematica obișnuită — și orice alt sistem formal particular care s-ar putea obține aplicând, de oricâte ori, lui $\mathbb{ZF}$, procedeul lui Gödel și orice alt sistem formal la care s-ar putea ajunge utilizând intuiția accesibilă matematicienilor, să zicem, în virtutea înțelegerii faptului că gödelizarea furnizează sisteme inatacabil corecte, sau alte tipuri de raționament sigur, de o natură mai puternică.

Algoritmul ar fi trebuit să conțină, ca instanțieri particulare ale sale, capacitatea de a face discriminări precise, deosebind argumentele valide de cele invalide în tot ceea ce urma să fie descoperit în activitatea matematică care în zilele noastre ocupă paginile revistelor de cercetare matematică. Acest presupus algoritm, incognoscibil, sau incomprehensibil ar fi trebuit să aibă codificat în interior puterea de a face toate acestea și totuși nouă ni se cere să credem că el a apărut doar prin selecția naturală pusă în mișcare de circumstanțele în care strămoșii noștri îndepărtați au luptat pentru supraviețuire.

Capacitatea specială de a face matematică nu putea să dea nici un avantaj direct, în selecție, posesorului său și eu cred că nu putea să existe nici un motiv pentru ca un astfel de algoritm să fi apărut.

Situația se schimbă atunci când presupunem că 'înțelegerea' este o calitate nealgoritmică. Atunci, ea nu mai trebuie să fie ceva atât de complicat încât să fie incognoscibilă sau incomprehensibilă. Într-adevăr, ea ar putea fi mult mai aproape de „ceea ce matematicienii cred că fac" Înțelegerea pare aparent că este o simplă calitate de bun simț. Este ceva greu de definit în mod limpede, dar fără îndoială ne este atât de familiară încât ne este greu să acceptăm că ar fi o calitate care să nu poată fi simulată corespunzător, măcar în principiu, de o procedură de calcul. Totuși, eu susțin chiar acest lucru. Dar pentru calcul este nevoie de o procedură algoritmică care să permită, pentru orice eventualitate, răspunsuri la toate întrebările din matematică, cu care vreodată s-ar putea cineva confrunta (într-un anumit sens), programate în algoritm. Dacă algoritmul nu este direct programat, atunci sunt necesare unele mijoace de calcul pentru găsirea unei metode de a da răspunsurile. Așa cum am văzut, aceste 'programe' sau 'metode de calcul' trebuie, dacă e să conțină tot ceea ce poate achiziționa înțelegerea umană, să fie ele însele deasupra înțelegerii umane.

Cum ar putea procesul orb al selecţiei naturale, pus în mişcare numai pentru a promova supravieţuirea strămoşilor noştri îndepărtaţi, să fie în stare să 'prevadă' că o procedură corectă de calcul, aşa şi aşa, care nu poate fi cunoscută, ar fi capabilă să rezolve probleme obscure de matematică care n-ar avea de fapt nici o importanţă pentru supravieţuire?

3.9. Algoritmi de învăţare

Ca nu cumva cititorul să fie prea uşor de acord că o astfel de posibilitate este absurdă, va trebui să fac mai clar tabloul pe care susţinătorii punctului de vedere computaţional sunt înclinaţi să îl prezinte. Aşa cum am văzut deja în paragraful 3.5, ei nu sunt preocupaţi de un algoritm care într-un sens să fi fost 'preprogramat' pentru a rezolva probleme de matematică, ci mai degrabă de anumite sisteme de calcul care să aibă capacitatea de *a învăţa*. Ei iau în considerare ceva care să aibă ingrediente *bottom-up* semnificative, împreună cu tot ceea ce procedurile *top-down* ar avea nevoie (vezi paragraful 1.5)[5]

Unii ar putea crede că descrierea *top-down* nu este deloc potrivită pentru un sistem care a apărut numai prin procesul orb al selecţiei naturale. Ceea ce înţeleg aici prin descriere *top-down* sunt acele aspecte ale presupusei proceduri algoritmice care sunt genetic fixe într-un organism şi nu sunt supuse modificărilor datorate învăţării din experienţă.

Deşi algoritmii *top-down* nu au fost proiectaţi cu nimic din 'cunoştinţele' actuale despre ceea ce ei ar fi trebuit să atingă în cele din urmă — ca secvenţe ADN care se traduc prin acţiuni corespunzătoare ale creierului — ei pot, fără îndoială, să furnizeze reguli clare prin care ar opera creierul în activitatea matematică. Aceste proceduri *top-down* vor furniza acele acţiuni algoritmice care constituie cadrul fix necesar în care 'procedurile de învăţare' (*bottom-up*), mai

[5] Există azi o teorie matematică a învăţării destul de bine elaborată; vezi Anthony şi Biggs (1992). Totuşi, această teorie se ocupă mai degrabă de probleme de complexitate, decât de probleme de calculabilitate — adică, de întrebări care privesc viteza şi spaţiul de memorie necesare pentru rezolvarea unor probleme; vezi MN p.156-163). Nu există nici o indicaţie că un astfel de sistem matematic de învăţare ar putea simula metoda prin care o persoană care se ocupă de matematică ajunge la noţiunea de 'adevăr indubitabil'

flexibile, vor fi lăsate să opereze.

Ne putem însă întreba: care este natura acestor proceduri de învățare ? Ne imaginăm că sistemul nostru de învățare este plasat într-un mediu extern, în care modul de acțiune al sistemului în acest mediu este modificat permanent după cum mediul a reacționat la acțiunile anterioare ale sistemului. Sunt aici, în principal, implicați doi factori. Factorul *extern* care este modul în care mediul se comportă și răspunde la acțiunile sistemului. Factorul *intern* este modul în care sistemul însuși își modifică comportarea ca răspuns la schimbările mediului său. Vom examina mai întâi natura algoritmică a factorului extern. Ar fi posibil ca reacția mediului exterior să furnizeze un ingredient nealgoritmic, chiar dacă construcția internă a sistemului de învățare este în întregime algoritmică?

În anumite condiții, așa cum este cazul cu 'antrenamentul' din rețelele neuronale artificiale, reacția mediului extern poate fi furnizată de comportarea unui experimentator, antrenor, sau profesor — să-i spunem cu un cuvânt 'profesor' — a cărui intenție este ca, în mod deliberat, să îmbunătățească performanțele sistemului. Atunci când sistemul se comportă într-un mod care este dorit de profesor, acest fapt este semnalat sistemului, astfel încât, în concordanță cu mecanismele sale interne de modificare a comportării proprii, sistemul să fie mai pregătit ca în viitor să se comporte în modul dorit de profesor.

De exemplu, să considerăm o rețea neuronală artificială antrenată pentru a recunoaște fețele oamenilor. Performanțele sistemului sunt continuu monitorizate, acuratețea cu care face 'ghicirile' este comunicată sistemului la fiecare pas, astfel încât el este capabil să-și îmbunătățească performanțele modificându-și în mod corespunzător structura internă. În practică, rezultatele acestor ghiciri nu trebuie monitorizate de un profesor adevărat, la fiecare etapă, deoarece procedura de antrenament poate fi automatizată. Dar în acest tip de situație, scopurile și judecățile persoanei profesorului formează criteriul ultim de performanță. În alte tipuri de situații, reacția mediului extern ar putea fi mai puțin 'deliberată' decât în exemplul descris mai sus. De exemplu, în cazul dezvoltării unui sistem *viu* — sistem imaginat a opera conform unei scheme de rețea neuronală (sau în cazul altei proceduri algoritmice, de exemplu, algoritmii genetici, așa ca în paragraful 3.7), la fel ca procedurile utilizate în modelele computaționale — nu sunt necesare astfel de evaluări externe ale performanțelor. Dimpotrivă, sistemul viu ar putea să-și modifice comportarea în conformitate cu criteriile *selecției naturale*, acționând

treptat după criterii care au funcţionat de-a lungul anilor şi care ar servi la creşterea propriilor şanse de supravieţuire şi la supravieţuirea progeniturilor sale.

3.10. Poate mediul să fie un factor extern nealgoritmic?

Acum suntem puşi în faţa unei situaţii în care sistemul însuşi (viu sau nu) trebuie să fie un fel de *robot* controlat de un calculator, deoarece procedurile de automodificare sunt în întregime computaţionale (folosesc termenul *robot* mai mult pentru a sublinia faptul că sistemul nostru trebuie văzut ca o entitate computaţională în interacţiune cu mediul. Prin aceasta nu presupun că ar fi vorba despre un dispozitiv mecanic construit în mod special de oameni. Ar putea să fie el însuşi o fiinţă umană în curs de dezvoltare, după principiile $\mathcal{A}$ sau $\mathcal{B}$, sau tot atât de bine, un obiect artificial). Ceea ce presupunem este că factorul *intern* este complet computaţional. Trebuie să ne întrebăm dacă factorul *extern* oferit de mediul înconjurător este sau nu computaţional — adică, trebuie să ne punem întrebarea dacă este posibilă simularea computaţională efectivă a mediului, atât în cazul *artificial* — când este comandat artificial de un profesor adevărat — sau în cazul *natural*, când forţele selecţiei naturale sunt arbitrul. În fiecare dintre aceste cazuri, regulile interne în conformitate cu care robotul, capabil să înveţe, îşi modifică propriul comportament, să poată răspunde semnalelor prin care mediul transmite modul în care au fost judecate performanţele anterioare.

Întrebarea dacă mediul poate fi simulat în cazul artificial, adică dacă un profesor în carne şi oase poate fi simulat computaţional, este de fapt pur şi simplu întrebarea pe care ne-am pus-o din nou şi din nou. În ipotezele $\mathcal{A}$ şi $\mathcal{B}$ a căror implicaţii le explorăm, se presupune că, în principiu, o simulare efectivă este posibilă. Până la urmă, plauzibilitatea acestei presupuneri trebuie investigată. Astfel, odată cu presupunerea că *robotul* nostru este computaţional, avem de asemenea un mediu computaţional. De aceea, sistemul *combinat*, constând din robot împreună cu mediul său de la care învaţă, ar trebui să fie ceva care, în principiu, să poată fi efectiv simulat computaţional, aşa încât mediul nu va oferi nici un fel de portiţă care să-i permită robotului computaţional să se comporte într-un mod non-computaţional.

Unii susţin că avantajul nostru faţă de calculatoare se datorează faptului că fiinţele umane formează o *comunitate*, în care, între membrii săi, comunicaţia este permanentă. În această viziune fiinţele umane pot fi considerate individual drept sisteme de calcul, dar comunitatea umană ar fi ceva mai mult. La fel ar putea fi privită şi comunitatea matematicienilor în raport cu indivizii care lucrează în domeniul matematicii — astfel, comunitatea ar putea acţiona într-un mod non-computaţional în vreme ce indivizii, nu. Îmi este imposibil să înţeleg această argumentaţie. Deoarece pot, la fel de bine, să consider o comunitate de calculatoare comunicând continuu între ele. O astfel de 'comunitate' va forma din nou un sistem computaţional, iar activitatea întregii comunităţi ar putea, la cerere, să fie simulată pe un singur calculator. Evident, comunitatea ar putea, pentru un mare număr de indivizi, să constituie un sistem computaţional imens, dar asta nu înseamnă o diferenţă de principiu.

Este adevărat că planeta noastră numără cam 5×10^9 locuitori (ca să nu menţionăm bibliotecile vaste conţinând cunoştinţe memorate). Dar aici nu este vorba decât de nişte numere mari şi, din punct de vedere computaţional, dezvoltarea calculatoarelor poate, în câteva decade, face faţă creşterii implicate de trecerea de la indivizi la comunitate. Pare limpede că, în cazul artificial, în care mediul exterior este constituit de persoana profesorului, nu obţinem nimic nou şi nu găsim nici o indicaţie cum ar putea să apară un element non-computaţional din aceşti constituenţi computaţionali.

Ce se poate spune în cazul natural? Întrebarea este dacă mediul *fizic*, considerat separat de persoana profesorului aflat în interiorul lui, ar putea conţine ingrediente care să nu poată fi, nici în principiu, simulate computaţional. Mi se pare că dacă cineva crede că există ceva ce în principiu nu poate fi simulat în mediul exterior, fără oameni, recunoaşte elementul principal care poate fi invocat împotriva lui $\mathcal{C}$. Deoarece singurul motiv limpede pentru a ne îndoi că $\mathcal{C}$ poate fi o posibilitate serioasă, se află în scepticismul cu care este privită eventualitatea că acţiunea unui obiect din lumea fizică poate acţiona în mod non-computaţional. Odată ce se admite că vreo acţiune fizică ar putea fi non-computaţională, apare posibilitatea ca acţiuni non-computaţionale să aibă loc şi în creierul fizic şi astfel este recunoscut principalul argument împotriva lui $\mathcal{C}$. În general, pare destul de improbabil ca să existe ceva în mediul exterior lipsit de oameni, care să eludeze calculul, decât ar fi în creierul uman (comparaţi de asemenea cu paragrafele 1.9

şi răspunsul la **Q2** din paragraful 2.6). Eu cred că s-ar putea susţine cu greu existenţa unui element relevant pentru mediul unui robot care învaţă şi care, în principiu, să fie non-computaţional.

Referindu-mă la posibilitatea de principiu ca mediul să fie de natură computaţională, trebuie să fac o precizare. Fără îndoială, mediul 'actual' al oricărui organism viu în dezvoltare (sau al unui robot sofisticat) depinde de factori incredibil de complicaţi şi o simulare a mediului, cu o precizie tolerabilă, nu poate fi imaginată. Chiar în cazul sistemelor fizice simple, comportarea dinamică poate fi extrem de complexă şi poate depinde atât de critic de detaliile stării iniţiale, încât nu se pune problema posibilităţii de a prevedea prin calcul comportarea lor următoare — prezicerea vremii pe termen lung este un exemplu de această natură, deseori citat. Astfel de sisteme sunt numite *haotice* (vezi şi paragraful 1.7) (sistemele haotice au o comportare elaborată şi imprevizibilă; aceste sisteme nu sunt, din punct de vedere matematic de neînţeles; ele sunt în prezent foarte studiate ca o parte importantă a cercetării matematice curente [8]). Aşa cum am arătat în paragraful 1.7, sistemele haotice se includ în ceea ce am numit 'calculabil' sau 'algoritmic' Aspectul esenţial al sistemelor haotice, din punctul nostru de vedere, este că nu e necesar să fie simulat vreun mediu haotic actual, ci simularea unui mediu *tipic* este suficientă. Astfel, de exemplu, nu trebuie să ştim timpul probabil, orice pronostic plauzibil este suficient.

3.11. Cum poate un robot să înveţe?

Să acceptăm că scopul nostru nu ar fi simularea computaţională a mediului extern. În ceea ce priveşte mediul, am putea să presupunem că nu există nici o piedică în simularea regulilor interne ale robotului însuşi.

Să ne punem deci întrebarea: cum învaţă robotul?

Ce fel de proces de învăţare este accesibil unui robot computaţional? S-ar părea că trebuie să existe reguli clare predefinite în cazul unor anumite sisteme artificiale de reţele neuronale, în general adoptate (vezi paragraful 1.5). În cazul acestor sisteme, trebuie să existe reguli de calcul bine definite, prin care legăturile dintre 'neuronii' artificiali care constituie reţeaua să fie întărite sau slăbite, astfel încât să îmbunătăţească performanţele în concordanţă cu criteriile (artificiale sau

naturale) stabilite de mediul exterior.

Un alt exemplu de sistem de învăţare este cel al *algoritmilor genetici*, în care funcţionează o anumită selecţie naturală între diferite proceduri algoritmice în funcţiune în maşină şi unde cel mai eficace algoritm de comandă a sistemului apare ca 'supravieţuirea celui mai bun'.

Aşa cum este normal, în cazul acestei organizări *bottom-up*, regulile diferă de regulile standard ale algoritmilor de calcul *top-down*, care lucrează în conformitate cu proceduri cunoscute pentru a furniza soluţii ale problemelor matematice. În loc de asta, aceste reguli *bottom-up* sunt pur şi simplu utilizate pentru ghidarea sistemului la modul general, pentru a-şi îmbunătăţi performanţele. Totuşi regulile trebuie să fie complet algoritmice — în sensul de a putea fi puse pe un calculator (maşină Turing).

În plus faţă de aceste reguli bine definite, în modul în care robotul nostru îşi modifică performanţele, ar trebui să fie încorporate şi elemente aleatoare. Aceste ingrediente aleatoare s-ar putea introduce într-un mod fizic, poate bazate pe procese cuantice ca, de exemplu, timpul de înjumătăţire în cazul atomilor radioactivi. În practică, tendinţa este de a utiliza în construcţia artificială a dispozitivelor, proceduri de calcul care să furnizeze la ieşire date aleatoare — numite pseudoaleatoare — chiar dacă ele sunt complet determinate de rezultatul unui calcul determinist (vezi paragraful 1.9). O altă procedură înrudită îndeaproape ar fi utilizarea timpului precis în care se face apel la o cantitate 'aleatoare', iar apoi acest timp să fie utilizat de un calcul complicat, de fapt un sistem haotic, în care chiar şi cele mai mici schimbări ale intrărilor duc la ieşiri impredictibile, efectiv aleatoare. Deşi ingredienţii aleatori ne scot în afara a ceea ce este acţiunea unei maşini Turing, ei nu fac asta în mod util.

O intrare pseudoaleatoare introdusă într-un robot, va fi, practic, echivalentă cu una aleatoare; iar data pseudoaleatoare nu ne scoate în afara a ceea ce face o maşină Turing.

Cititorul ar putea în acest punct să se neliniştească în faţa faptului că deşi *în practică* o dată aleatoare nu ar fi diferită de una pseudoaleatoare, există totuşi o diferenţă *de principiu*. Ne-am preocupat şi până acum de distincţia dintre ceea ce poate fi obţinut de către matematicieni 'în principiu' şi ceea ce poate fi obţinut 'în practică' (vezi şi paragrafele 3.2—3.4).

Există anumite tipuri de situaţii matematice în care *de fapt* datele aleatoare

furnizează o soluţie la o problemă, când, tehnic vorbind, nici o dată pseudoaleatoare nu reuşeşte acest lucru. O astfel de situaţie apare când problema implică un element 'competitiv', ca în teoria jocurilor şi în criptografie. Pentru anumite tipuri de 'jocuri cu doi jucători', strategia optimă implică un ingredient complet aleator [9]. Orice abatere semnificativă, din partea unuia dintre jucători, de la strategia aleatoare care este optimală, ar permite, pentru o secvenţă suficient de lungă de partide, celuilalt jucător să obţină un avantaj — cel puţin în principiu. Acest avantaj ar putea să intervină dacă, într-un mod oarecare, oponentul ar putea să ghicească destul de bine natura pseudoaleatoare a ingredientului pe care primul jucător l-a utilizat

O situaţie similară apare în criptografie unde securitatea unui cod depinde de şirul de cifre generat în mod pur aleator. Dacă şirul nu este generat aleator ci cu un proces pseudoaleator, atunci, din nou, există o şansă ca natura detaliată a procesului pseudoaleator să fie identificat de cel care încearcă să spargă codul — dacă reuşeşte, va intra în posesia unei informaţii de nepreţuit.

La prima vedere, ar putea să pară că dacă în astfel de situaţii de competiţie, caracterul aleator este preţios, atunci el ar putea fi favorabil şi în selecţia naturală. Într-adevăr, eu sunt sigur că este un factor important în dezvoltarea organismelor, în multe privinţe. Totuşi, aşa cum vom vedea mai târziu în acest capitol, caracterul aleator pur şi simplu nu ne permite să scăpăm din plasa lui Gödel. Chiar dacă ingredienţii pur aleatori pot fi trataţi ca parte a argumentaţiei care urmează, ei nu ne permit să evadăm din constrângerile pe care trebuie să le îndeplinească sistemele de calcul. De fapt, există o sferă de acţiune ceva mai mare în cazul proceselor pseudoaleatoare, decât în cazul celor aleatoare (aşa cum rezultă şi din paragraful 3.22).

Pentru moment, să presupunem că sistemul nostru robot este de fapt o maşină Turing (cu o capacitate finită de memorie). Mai exact, cum robotul este permanent în interacţiune cu mediul său şi noi am presupus că mediul poate de asemenea să fie simulat computaţional, va trebui să considerăm că robotul, împreună cu mediul său, acţionează ca o singură maşină Turing. Totuşi, va fi util să considerăm robotul separat, ca o maşină Turing şi să privim mediul ca furnizând informaţii pe banda de intrare a maşinii. De fapt, această analogie nu este prea potrivită căci, din motive tehnice, maşina Turing este un lucru *fix* despre care nu se presupune că îşi modifică structura pe baza 'experienţei'

Am putea să încercăm să ne imaginăm că maşina Turing şi-ar schimba structura pe măsură ce funcţionează şi informaţiile din mediul exterior sunt continuu introduse în maşină de pe banda de intrare. Totuşi nu se întâmplă aşa, pentru că ieşirea maşinii Turing nu se presupune ca ar fi examinată până când maşina nu atinge comanda sa internă STOP, punct în care nu se presupune că ar mai examina banda de intrare decât dacă funcţionarea porneşte din nou (vezi paragraful 2.1 şi anexa A; de asemenea **MN** cap. 2). Pentru o rulare următoare a maşinii, ea trebuie să revină la starea sa originară şi, în acest mod, nu poate 'învăţa'

Totuşi, această dificultate poate fi remediată uşor prin următorul dispozitiv tehnic. Luăm maşina noastră Turing ca fixă, dar după fiecare citire a benzii să presupunem că maşina scoate *două* informaţii (codificate printr-un singur număr) atunci când atinge comanda STOP. Primul lucru ar codifica comportarea sa externă calculată, în vreme ce al doilea ar fi de uz intern, codificând toată experienţa pe care a obţinut-o din interacţiunile anterioare cu mediul extern. La următoarea rulare, maşina citeşte mai întâi aceste informaţii 'interne' de pe banda de intrare, iar apoi, ca o a doua parte, de pe banda de intrare, toate informaţiile 'externe' pe care mediul i le furnizează, incluzând reacţia detaliată pe care mediul a avut-o faţă de comportarea anterioară a maşinii. Astfel, tot ceea ce învaţă este codificat în partea internă a informaţiei de pe bandă şi maşina se alimentează cu această informaţie de pe bandă, introducând-o din nou în maşină (deşi banda va fi tot mai lungă cu timpul).

3.12. Poate un robot să aibă 'convingeri matematice ferme'?

În modul discutat în paragrafele precedente, putem descrie cel mai general 'robot' care învaţă ca pe o maşină Turing. Despre robotul nostru se presupune că este în stare să emită judecăţi matematice adevărate, având capacitatea potenţială a oamenilor care fac matematică. Cum se realizează acest lucru ? Nu vrem să pretindem ca toate regulile matematice (ca acelea pe care le conţine un sistem formal ZF) să fie codificate în întregime într-un mod *top-down*, aşa cum ar trebui dacă am vrea să cuprindem direct intuiţiile la care pot ajunge matematicienii,

deoarece, aşa cum am văzut, nu există un mod rezonabil (o prescurtare pentru 'intervenţia divină', conform cu paragrafele 3.5, 3.6) pentru ca un astfel de algoritm *top-down*, vast, complicat şi incognoscibil să poată fi implementat.

Trebuie să presupunem că ori ce fel de elemente *top-down* ar exista, acestea nu sunt potrivite pentru a face matematică sofisticată, ci sunt reguli generale care ar putea fi imaginate ca furnizând o bază pentru *înţelegere*.

Să ne amintim cele două tipuri de date de intrare provenind din mediul extern, aşa cum le-am menţionat mai sus, (vezi paragraful 3.9) şi care pot influenţa semnificativ comportarea robotului nostru: cele artificiale şi cele naturale. Pentru intervenţia artificială a mediului ne-am imaginat un profesor (sau mai mulţi) care instruiesc robotul în legătură cu diferitele adevăruri matematice şi încearcă să-l conducă astfel încât să ajungă să găsească un mod intern de a distinge singur adevărul de fals. Profesorul poate să informeze robotul atunci când face greşeli, sau îi poate spune despre diferite concepte matematice şi metode acceptabile de demonstraţie matematică. Procedeele aplicate de profesor ar putea fi de diverse feluri, începând cu învăţarea prin 'exemple', prin 'ghidare', prin 'instruire' sau chiar 'vorbindu-i'! În ceea ce priveşte aspectele naturale ale mediului fizic, acestea pot furniza robotului 'idei' despre comportarea obiectelor fizice; mediul poate furniza şi realizări concrete ale conceptelor matematice, ca de exemplu diferite instanţieri ale numerelor naturale: două mere, şapte portocale, patru banane, zero pantofi, un pantalon, etc. şi aproximări bune ale ideilor geometrice ca, de exemplu, dreapta, cercul şi de asemenea aproximări ale noţiunii de mulţime infinită (ca de exemplu mulţimea punctelor situate în interiorul unui cerc).

Deoarece robotul nostru nu este programat complet în modul *top-down* şi se presupune că ajunge la conceptul de adevăr matematic cu ajutorul procedurii de învăţare, va trebui să-i permitem să facă greşeli, ca parte din activitatea sa de învăţare astfel încât să înveţe din greşeli. La început aceste greşeli pot fi corectate de profesorii săi. Uneori, ca o alternativă, robotul poate observa din mediul extern că unele dintre sugestiile asupra adevărului matematic sunt eronate, sau par eronate. Sau poate ajunge la astfel de concluzii pe baza unor consideraţii interne asupra consistenţei, etc. Ideea ar fi că robotul va face tot mai puţine erori, pe măsură ce experienţa sa creşte. Pe măsura trecerii timpului, profesorii şi mediul fizic înconjurător devin tot mai puţin importanţi şi probabil devin eventual chiar inoportuni pentru formarea judecăţilor sale matematice, iar robotul poate să se

bazeze tot mai mult pe puterea sa internă de calcul. Astfel, se poate presupune că robotul nostru ar putea să depăşească acele adevăruri matematice pe care le-a învăţat de la profesorii săi sau le-a dedus din experienţa proprie în contact cu mediul fizic. Se poate astfel imagina că va aduce contribuţii originale la cerecetarea matematică.

Pentru a examina plauzibilitatea acestor fapte, va trebui să punem aceste lucruri în legătură cu ceea ce am discutat mai devreme. Dacă robotul nostru are cu adevărat capacitatea, înţelegerea şi intuiţia unui matematician adevărat, i se cere un fel anume de concept — acela al 'adevărului matematic de necontestat' Încercările sale mai vechi care fuseseră corectate de profesori, sau dovedite ca neplauzibile de mediul înconjurător, nu intră în această categorie. Acelea intră în categoria lucrurilor 'ghicite', ghicitul fiind exploratoriu şi permiţând erori. Dacă robotul este capabil să se comporte ca un matematician pur, el va face, din timp în timp, greşeli — în principiu greşeli corectabile, în conformitate cu propriul criteriu de 'adevăr indubitabil'.

Am văzut din discuţia purtată anterior că un concept uman de 'adevăr incontestabil' nu poate fi atins de nici o mulţime de reguli mecanice care să poată fi cunoscute de oameni . Dacă acum presupunem că robotul nostru este capabil să atingă (sau chiar să întreacă) nivelul capacităţilor matematice pe care o fiinţă omenească poate să-l atingă, în principiu, atunci conceptul *său* de adevăr matematic indubitabil trebuie să fie de asemenea imposibil de cuprins de vreo mulţime de reguli mecanice care, în principiu, ar putea fi percepute drept corecte — acest lucru înseamnă aici percepute de către un matematician profesionist sau, în aceste condiţii, de robotul nostru matematician!

Întrebarea importantă ar fi deci: ale cui concepte, percepţii sau convingeri de nezdruncinat trebuie considerate relevante — ale noastre sau ale robotului?

Poate fi un robot considerat ca având de fapt percepţii şi convingeri? Cititorul, dacă este un susţinător al punctului de vedere $\mathcal{B}$, ar putea avea dificultăţi cu această întrebare deoarece conceptele *percepţie* şi *convingere* sunt atribute *mentale* şi ar trebui privite ca neaplicabile unui robot complet comandat de un calculator. Totuşi, în discuţia de mai sus, nu este cu adevărat necesar ca robotul să posede calităţi mentale pure, cu condiţia că el să se comporte *extern* exact cum s-ar comporta un matematician uman, de fapt, aşa cum ar implica aderenţa strictă la $\mathcal{B}$ şi la $\mathcal{A}$. Astfel, nu este necesar ca robotul să înţeleagă, să perceapă sau să

fie convins de ceva, cu condiția ca, din punctul de vedere al comportamentului extern, acesta să dea impresia că robotul posedă aceste capacități mentale. Voi relua aceste idei mai amănunțit în paragraful 3.17.

Punctul de vedere $\mathcal{B}$ nu diferă, în principiu, de $\mathcal{A}$, în ceea ce privește limitările modului în care un robot poate să se comporte, dar susținătorii lui $\mathcal{B}$ ar putea să aibă pretenții mai modeste cu privire la ceea ce poate de fapt atinge un robot, sau despre posibilitatea de a găsi un sistem computațional care să poată fi considerat că este în stare să realizeze o simulare efectivă a creierului unei persoane care se află în procesul de percepere a validității unei demonstrații matematice. O astfel de percepție umană implică o înțelegere a *semnificațiilor* conceptelor matematice utilizate.

Potrivit punctului de vedere $\mathcal{A}$, nu există nimic situat deasupra unor caracteristici ale calculului care ar fi implicate în chiar noțiunea de *semnificație*, în vreme ce, în conformitate cu $\mathcal{B}$, *semnificația* este un aspect semantic al activității mentale și diferă de orice ar putea fi descris în termeni pur computaționali.

În concordanță cu $\mathcal{B}$, nu putem să ne așteptăm ca robotul nostru să poată face aprecieri semantice de nici un fel. Astfel, suporterii punctului de vedere $\mathcal{B}$, sunt mai puțin înclinați, decât cei ai lui $\mathcal{A}$, să creadă că orice robot, construit după principiile pe care le-am considerat, ar putea de fapt atinge manifestările externe ale înțelegerii umane pe care un matematician obișnuit este capabil să le atingă. Eu cred că acest lucru sugerează (nu în mod nenatural) că suporterii lui $\mathcal{B}$ pot fi mai ușor convertiți la $\mathcal{C}$ decât ar putea fi convertiți suporterii lui $\mathcal{A}$; dar, din punctul de vedere a ceea ce dorim să stabilim aici, diferențele dintre $\mathcal{A}$ și $\mathcal{B}$ sunt nesemnificative.

Din toate acestea se desprinde concluzia că, deși afirmațiile matematice ale robotului nostru, controlate de un sistem de proceduri de calcul *bottom-up* sunt exploratorii și oarecum provizorii în ceea ce privește adevărul lor, va trebui să presupunem că robotul mai are un nivel de securitate în plus, al 'convingerilor' matematice de neatacat, astfel încât unele dintre aserțiunile sale — a căror atestare va fi marcată de un simbol special, să zicem '★', să fie considerate de neatacat în conformitate cu criteriile *proprii* ale robotului.

Întrebarea dacă robotului i se permite să facă erori de notare a aserțiunilor cu ★ — eventual corectabile de robotul însuși — va fi examinată în paragraful 3.19. Pentru moment, vom presupune că, îndată ce robotul a marcat o aserțiune

cu ★, aceasta o considerăm cu adevărat fără de erori.

3.13. Mecanisme care stau la baza matematicii robotului

Să considerăm acum toate mecanismele care intră în procedurile care guvernează comportarea robotului, astfel încât, la sfârşit să obţină aserţiunile notate cu ★. Unele dintre acestea vor fi *interne*, ale robotului însuşi.

Ar trebui să existe şi anumite proceduri prin care robotul să-şi poată îmbunătăţi performanţele (şi să progreseze gradat până la nivelul ★). Toate acestea ar trebui să fie în mod normal considerate ca lucruri care pot fi cunoscute omului (chiar dacă implicaţiile finale ale tuturor acestor factori împreună ar putea depăşi capacităţile de calcul ale unui matematician uman).

Evident, dacă ar fi să presupunem că fiinţele umane vor putea, într-o zi, să construiască un robot în stare să facă matematică pură, atunci mecanismul intern pe baza căruia este construit de fapt robotul trebuie să poată fi cunoscut de un om; altfel încercarea de a construi un astfel de robot ar fi zadarnică!

Bineînţeles, va trebui să permitem ca procesul de construcţie să aibă mai mulţi paşi; adică, robotul nostru, în stare să facă matematică, să fie construit în întregime de roboţi 'de ordin inferior' (aceia să nu fie, de fapt, capabili să facă matematică pură) şi probabil acei roboţi să fi fost şi ei construiţi, la rândul lor, de roboţi de un nivel încă mai coborât. Totuşi, ierarhia întreagă ar fi trebuit să fi fost pusă în mişcare de o fiinţă umană, iar regulile de funcţionare (posibil un amestec de proceduri *top-down* şi *bottom-up*) să poată fi cunoscute de oameni.

Trebuie, de asemenea, să includem, ca ingredient esenţial al dezvoltării robotului, toţi factorii *externi* venind din mediul său extern. Pot într-adevăr să existe o mulţime de intrări provenind din mediu, atât de la profesorul uman (sau robot), cât şi de la mediul fizic natural. Aceşti factori 'naturali' provenind de la mediul neuman, nu ar trebui să-i considerăm 'incognoscibili'. Aceştia ar putea fi foarte complicaţi, interactivi, dar există deja 'realitatea virtuală' efectivă care simulează aspectele semnificative ale mediului înconjurător (ca în paragraful 1.20). Pare să nu existe nici un motiv pentru ca această simulare să nu poată fi extinsă până la a oferi robotului nostru tot de ceea ce ar avea nevoie pentru a se dezvolta,

în modul dictat de factorii naturali externi — să nu uităm (vezi paragrafele 1.7, 1.9) — că tot ceea ce trebuie simulat este un mediu *tipic*, nu neapărat un mediu particular dat.

Intervenţia umană (sau a unui robot) — care constituie aici factorii externi 'artificiali' — poate să aibă loc în stadii foarte diferite, dar acest lucru nu schimbă în mod esenţial aspectul cognoscibil al mecanismelor de bază, cu condiţia ca intervenţia umană să fie ceva cognoscibil, care să poată fi mecanizat. O astfel de presupunere este oare echitabilă? Sunt sigur că ar putea părea natural — mai cu seamă pentru suporterii lui $\mathcal{A}$ şi $\mathcal{B}$ — să presupună că intervenţia umană în dezvoltarea unui robot ar putea fi înlocuită de una complet computaţională. Noi nu cerem ca această intervenţie să fie ceva misterios — să zicem, ceva de 'esenţă' indefinibilă pe care profesorul ar putea să o furnizeze robotului, ca parte a educaţiei sale. Ne aşteptăm doar ca să existe unele informaţii de bază care trebuie să fie furnizate robotului, pe care le poate da mai uşor o fiinţă umană. Este foarte probabil ca, în cazul în care unui elev i se predă ceva, furnizarea optimă a informaţiei să se realizeze în mod interactiv, astfel încât comportamentul profesorului să depindă de modul în care reacţionează elevul. Dar acest lucru, nu este în sine un motiv de a contesta posibilitatea ca rolul profesorului să nu fie unul complet computaţional.

Întreaga discuţie este, în acest capitol, una *de reducere la absurd* în care presupunem că nu există nimic esenţial non-computaţional în comportarea unei fiinţe umane. Pentru cei care susţin punctele de vedere $\mathcal{C}$ sau $\mathcal{D}$, care sunt mai degrabă dispuşi să creadă în posibilitatea unei 'esenţe' non-computabile, furnizată robotului în virtutea umanităţii profesorului său, această întreagă discuţie nu este în nici un caz necesară!

Punând toate aceste mecanisme împreună (cele care constau în proceduri interne de calcul şi cele datorate interacţiunii cu mediul extern), nu pare să fie rezonabil să le considerăm ca ceva ce nu poate fi cunoscut, deşi unii ar putea adopta opinia că toate implicaţiile acestor mecanisme externe nu ar fi omeneşte calculabile — sau că nu ar fi calculabile în practică, de nici un calculator actual sau viitor. Mă întorc, însă pe scurt, la problema cognoscibilităţii tuturor acestor mecanisme computaţionale (vezi sfârşitul paragrafului 3.15). Să presupunem deocamdată că aceste mecanisme ar putea fi cunoscute — şi să notăm această mulţime de mecanisme cu M. S-ar putea întâmpla ca unele aserţiuni-$\bigstar$, la care

conduc aceste mecanisme, să nu poată fi cunoscute de oameni? Este acesta un punct de vedere consistent? Nu este, dacă vom continua să interpretăm aici 'cognoscibilitatea' în sensul de posibilitate 'principială' de cunoaştere, aşa cum am adoptat-o în legătură cu cazurile **I** şi **II** şi aşa cum a fost enunţată explicit la începutul paragrafului 3.5. Faptul că ceva (de exemplu, formularea unor aserţiuni ★) s-ar putea afla dincolo de puterea de calcul a unui om, lipsit de calculator, nu are nici o semnificaţie aici. Mai mult, nu am avea nici o obiecţie să permitem în procesul de învăţare să se utilizeze creionul şi hârtia, sau un calculator de buzunar, sau chiar un calculator programat *top-down*.

Includerea ingredienţilor *bottom-up* în procedurile de calcul nu aduce nimic nou în ceea ce poate fi atins în principiu — cu condiţia ca mecanismele de bază implicate în aceste proceduri *bottom-up* să poată fi înţelese. Pe de altă parte, 'cognoscibilitatea' mecanismelor **M** trebuie luată într-un sens 'practic', aşa cum am considerat explicit în paragraful 3.5. Deci, pentru moment, vom presupune că mecanismele **M** pot fi cu adevărat cunoscute *în practică*.

Cunoscând mecanismele **M**, le putem lua drept bază pentru construirea unui sistem formal $\mathbb{Q}(\mathbf{M})$ unde teoremele din $\mathbb{Q}(\mathbf{M})$ vor fi:

(i) aserţiunile-★ care apar din implementarea acestor mecanisme;
(ii) orice propoziţie care poate fi obţinută din aceste aserţiuni-★ utilizând regulile logicii elementare.

Prin 'logică elementară' vom înţelege, să zicem, regulile *calculului predicatelor* — în acord cu ceea ce am spus în paragraful 2.9 — sau alt sistem de reguli logice, (calculabile), clar definite şi de necontestat. Noi putem cu adevărat să construim sistemul formal $\mathbb{Q}(\mathbf{M})$ în virtutea faptului că există o procedură $Q(\mathbf{M})$ (deşi poate una foarte lungă) pe baza căreia se pot obţine unele după altele aserţiunile-★, din **M**. Să observăm că $Q(\mathbf{M})$, aşa cum a fost definit, generează aserţiunile din (i) de mai sus, dar nu şi pe cele din (ii), cu necesitate (deoarece, putem presupune că robotul ar considera foarte plicticoasă simpla generare a implicaţiilor logice ale teoremelor-★ pe care le produce!). Astfel, cu siguranţă $Q(\mathbf{M})$ este diferit de $\mathbb{Q}(\mathbf{M})$, dar diferenţa nu este importantă. Bineînţeles, dacă dorim, putem extinde procedura de calcul $Q(\mathbf{M})$ pentru a obţine alta care *să fie* echivalentă cu $\mathbb{Q}(\mathbf{M})$.

Pentru interpretarea sistemului formal $\mathbb{Q}(\mathbf{M})$, trebuie să fie clar că pe

măsură ce robotul se dezvoltă, semnul tipografic ★ înseamnă — şi va continua să însemne — că ceea ce se afirmă trebuie luat ca stabilit fără dubii. În lipsa instruirii de către profesorii umani, nu putem fi siguri că robotul nu-şi dezvoltă, pentru uzul propriu un limbaj diferit, în care semnul ★ să aibă o cu totul altă semnificaţie, dacă are vreo semnificaţie. Pentru a ne asigura că limbajul robotului este consistent cu specificaţiile noastre din definiţia lui $\mathbb{Q}(\mathbf{M})$, trebuie să fim siguri că, în antrenamentul robotului (să zicem, făcut de un profesor uman) semnificaţia care trebuie ataşată semnului ★ este cea dorită. La fel, trebuie să putem fi siguri că notaţia pe care robotul o utilizează pentru ca să specifice, să zicem, propriile propoziţii π_1, este aceeaşi (sau translatabilă în mod explicit) cu notaţiile pe care le folosim noi înşine. Dacă mecanismul **M** poate fi cunoscut de oameni, rezultă că axiomele şi regulile de procedură ale sistemului formal $\mathbb{Q}(\mathbf{M})$ trebuie să fie cognoscibile. Mai mult, orice teoremă obţinută în interiorul sistemului $\mathbb{Q}(\mathbf{M})$ poate fi considerată, în principiu, drept cognoscibilă (în sensul că specificaţiile sale pot fi cunoscute de un om, chiar dacă nu şi adevărul lor), chiar dacă procedurile de calcul pentru obţinerea multor teoreme de acest fel ar putea să depăşească posibilităţile de calcul ale unui om, neasistat de calculator.

3.14. Contradicţia fundamentală

Ceea ce a reuşit discuţia de mai sus să spună este faptul că „algoritmul necunoscut şi inconştient F ", pe care **III** îl consideră că stă la baza percepţiei adevărului matematic, poate fi redus la unul conştient cognoscibil — cu condiţia ca, în conformitate cu scopurile **IA**, să poată fi posibilă construcţia unui robot capabil să facă matematică la nivel uman (sau peste). Algoritmul incognoscibil F este astfel înlocuit cu sistemul formal cognoscibil $\mathbb{Q}(\mathbf{M})$.

Înainte de a examina aceste lucruri în detaliu, aş dori să vă atrag atenţia asupra unui lucru pe care nu l-am presupus încă şi anume posibilitatea de a fi introduse, la diferite nivele ale dezvoltării roboţilor, a unor elemente *aleatoare*, în plus faţă de mulţimea fixă de mecanisme. Acestui lucru va trebui să-i dăm atenţia cuvenită, în cele ce urmează, dar pentru moment, voi considera că aceste elemente aleatoare pot fi obţinute prin calcul (haotic) pseudoaleator. Aşa cum am spus în paragrafele 1.9, 3.11, aceste ingrediente pseudoaleatoare ar putea fi potrivite, în

practică. Mă voi întoarce la problema intrărilor aleatoare în paragraful 3.18, unde se va discuta mai complet aspectul pur aleator, dar deocamdată, când mă refer la mecanismele **M**, voi presupune că ele sunt pe deplin calculabile şi lipsite de incertitudini.

Ideea centrală a contradicţiei noastre este, în mare, faptul că $\mathcal{Q}(\mathbf{M})$ trebuie să ia locul lui F, din discuţia noastră anterioară, cea dată în paragraful 3.2 în legătură cu cazul **I**. Deoarece cazul **III** a fost efectiv redus la cazul **I** şi prin aceasta a fost epuizat.

Vom presupune – în conformitate cu $\mathcal{A}$ şi $\mathcal{B}$, de dragul argumentaţiei – că, în principiu, robotul nostru ar putea, prin procesul de învăţare de natura celui pe care l-am descris deja, să obţină orice rezultat matematic dintre cele pe care un om le poate obţine. Trebuie să acceptăm că ar putea obţine rezultate care, în principiu, întrec puterea omenească. În orice caz, robotul ar putea aprecia forţa argumentaţiei lui Gödel (sau ar putea cel puţin să simuleze această apreciere, în concordanţă cu $\mathcal{B}$). Astfel, pentru orice sistem formal dat (suficient de extins), $\mathbb{H}$, robotul ar trebui să fie capabil să perceapă indubitabil faptul că corectitudinea lui $\mathbb{H}$ implică adevărul propoziţiei[6] Gödel, corespunzătoare, $G(\mathbb{H})$ şi că implică de asemenea că $G(\mathbb{H})$ nu este o teoremă în $\mathbb{H}$ În particular, robotul ar percepe că adevărul lui $G(\mathcal{Q}(\mathbf{M}))$ urmează fără dubii din corectitudinea lui $\mathcal{Q}(\mathbf{M})$ şi de asemenea faptul că $G(\mathcal{Q}(\mathbf{M}))$ nu este o teoremă în $\mathcal{Q}(\mathbf{M})$.

Aşa cum a rezultat pentru cazul **I** (discutat în paragraful 3.2, pentru fiinţele umane) rezultă de aici că robotul este incapabil să creadă cu fermitate că sistemul formal $\mathcal{Q}(\mathbf{M})$ este echivalent cu propriul său concept de convingere matematică indubitabilă. Asta în ciuda faptului că noi (adică experţii **IA**), am putea şti foarte bine că mecanismul **M** *stă* la baza sistemului de convingeri matematice şi deci că sistemul de convingeri indubitabile este echivalent cu $\mathcal{Q}(\mathbf{M})$. Deoarece, dacă robotul crede cu fermitate că sistemul propriilor sale convingeri ar fi încapsulate în $\mathcal{Q}(\mathbf{M})$, atunci ar trebui să creadă în corectitudinea lui $\mathcal{Q}(\mathbf{M})$. În consecinţă, ar trebui să creadă în $G(\mathcal{Q}(\mathbf{M}))$ şi, în acelaşi timp, să considere că $G(\mathcal{Q}(\mathbf{M}))$ se află în afara sistemului propriu de convingeri ceea ce este o contradicţie! Astfel,

---

[6] Într-o primă versiune a acestei cărţi (ediţia Oxford University Press Inc., New York (n.t.)), în cele ce urmează din capitolul 3, s-a folosit $\Omega(\mathbb{F})$ în locul lui $G(\mathbb{F})$. Utilizarea lui $G(\mathbb{F})$ este mai potrivită totuşi, (vezi şi paragraful 2.8).

robotul ar fi incapabil să știe că el a fost construit în conformitate cu mecanismele **M**. Cum noi suntem conștienți — sau cel puțin ar trebui să fim conștienți — că robotul a fost construit așa, acest lucru pare că ne spune că noi avem acces la adevărul matematic, adică la $G(\mathbb{Q}(\mathbf{M}))$, lucru care se află deasupra posibilităților robotului, în ciuda faptului că abilitățile lui sunt considerate a fi egale cu (sau mai mari decât) capacitățile umane.

3.15. Mijloace prin care contradicția poate fi îndepărtată

Putem considera argumentația în două moduri diferite. Putem să o privim din punctul de vedere al creatorilor umani ai robotului, sau din punctul de vedere al robotului. Din perspectivă umană, pretențiile robotului la un adevăr de neatacat nu sunt în mod special convingătoare pentru un matematician uman, în afară de cazul în care argumentele actuale individuale pe care robotul le folosește, sunt de apreciat în sine.

Teoremele din $\mathbb{Q}(\mathbf{M})$ s-ar putea să nu fie acceptate toate de om ca incontestabil adevărate — și vă reamintim că puterea de a raționa a robotului ar putea fi de fapt mai presus de capacitatea umană. Astfel, s-ar putea argumenta că numai cunoașterea faptului că robotul a fost construit după mecanismele **M** nu poate constitui o demonstrație matematică de necombătut (de către oameni). Ca atare, vom lua toată argumentația ca fiind prezentată din punctul de vedere al robotului. Să vedem ce fel de lacune s-ar putea găsi în argumentația pe care robotul ar putea-o face.

S-ar părea că robotul are la dispoziție doar patru posibilități pentru a împiedica această contradicție presupunând că acceptăm că există un asemenea fel de robot computațional.

(a) Acceptând că **M** ar putea sta la baza propriei construcții, robotul nu va putea fi totuși convins indubitabil de acest fapt.

(b) Fiind convins indubitabil de fiecare aserțiune-★ în momentul în care o face, robotul ar putea totuși să aibă îndoieli că *întregul* sistem de aserțiuni-★ ar fi de încredere ca atare, ar putea să nu fie convins că $\mathbb{Q}(\mathbf{M})$ stă, de fapt, la baza sistemului său de aprecieri cu privire la propozițiile π_1.

(c) S-ar putea ca mecanismul adevărat **M** să depindă în mod esențial de elemente *aleatoare* și să nu poată fi descris în termenii unor elemente pseudo-aleatoare
(d) S-ar putea ca adevăratul mecanism **M** să nu poată de fapt fi cunoscut.

Scopul următoarelor nouă paragrafe este să prezinte cu detalii argumentele care arată că nici una dintre variantele (a), (b) și (c) nu oferă o metodă plauzibilă de a scoate robotul din această contradicție. Ca urmare și noi și robotul, suntem conduși spre concluzia neplăcută (d), dacă rămânem să insistăm că înțelegerea matematică poate fi redusă la calcul. Sunt sigur că specialiștii în inteligență artificială vor fi tot atât de nemulțumiți ca și mine de varianta (d). Situația oferă probabil (așa cum am spus în paragraful 1.3, la sfârșit, prezentând sugestia $\mathcal{A}/\mathcal{D}$) un punct de sprijin prin care *intervenția divină* ar fi necesară pentru implantarea unui algoritm necunoscut în toate creierele calculatoarelor (de către Cel Dintâi Programator din branșă). Oricum, concluzia 'incognoscibil' pentru însuși mecanismul responsabil de inteligența noastră — n-ar fi o concluzie prea fericită, nici chiar pentru aceia dintre noi care sperăm să înțelegem, în principiu, cum a apărut inteligența umană, utilizând legi științifice comprehensibile, ca acelea din fizică, chimie, biologie, selecție naturală — fără a dori să reproducem o astfel de inteligență într-un robot. În opinia mea, o astfel de concluzie pesimistă nu este justificată deoarece 'comprehensibilitatea' științifică este un lucru foarte diferit de calculabilitate. Concluzia va fi că nu legile de bază ar fi incomprehensibile, ci că ele sunt non-computaționale. Despre această problemă va trebui să vorbesc pe larg în partea a doua a acestei cărți.

3.16. Trebuie robotul să creadă în M ?

Să ne imaginăm că prezentăm robotului o mulțime **M** de mecanisme — care ar putea fi cele care stau la baza construcției sale. Voi încerca să-l conving pe cititor că robotul va respinge posibilitatea ca **M** să stea la baza capacității sale de a înțelege matematica — indiferent dacă acest lucru este așa, sau nu! Să presupunem, pentru moment, că robotul respinge posibilitățile (b), (c) și (d), așa încât să ajungem la concluzia oarecum surprinzătoare ca (a) nu poate, singur, să ne permită să scăpăm de paradox.

Raționamentul este următorul. Fie $\mathcal{M}$ ipoteza:

„mecanismul M stă la baza înţelegerii matematice"

Să considerăm acum aserţiuni de forma:

„o propoziţie π_1 (cu proprietăţile), este o consecinţă a lui $\mathcal{M}$"

Voi numi o astfel de aserţiune, dacă robotul crede cu fermitate în ea, o 'aserţiune $\bigstar_{\mathcal{M}}$' Astfel aserţiunile $\bigstar_{\mathcal{M}}$ nu se referă în mod necesar la propoziţiile-π_1 care în sine sunt crezute indubitabil de robot, ci şi la propoziţiile-π_1 despre care robotul acceptă că sunt deducţii inatacabile din ipoteza $\mathcal{M}$. Robotul nu trebuie, pentru început să aibă nici un fel de părere despre modul în care a fost el construit (în concordanţă cu **M**). S-ar putea ca, iniţial, să fie de părere că această variantă este improbabilă, dar, fără îndoială, va putea să analizeze perfect — în cea mai autentică tradiţie ştiinţifică — care sunt consecinţele indubitabile ale ipotezei după care a fost construit.

Există atunci propoziţii-π_1 pe care robotul să le considere drept consecinţe indubitabile ale lui $\mathcal{M}$ şi care, în plus, nu sunt simple aserţiuni-$\bigstar$ care ar avea nevoie de $\mathcal{M}$? Într-adevăr, există. Deoarece, aşa cum am observat la sfârşitul paragrafului 3.14, adevărul propoziţiei-π_1, $G(\mathbb{Q}(\mathbf{M}))$ rezultă din corectitudinea lui $\mathbb{Q}(\mathbf{M})$ şi tot aşa rezultă că $G(\mathbb{Q}(\mathbf{M}))$ nu este o teoremă în $\mathbb{Q}(\mathbf{M})$. Mai mult, robotul va fi indubitabil convins de această implicaţie. Presupunând că robotul este de acord cu faptul că indubitabilele sale convingeri sunt încapsulate de $\mathbb{Q}(\mathbf{M})$, dacă el însuşi este construit pe baza lui **M** (adică respingând posibilitatea[7] (b)) şi că atunci corectitudinea lui $\mathbb{Q}(\mathbf{M})$ este o consecinţă a lui $\mathcal{M}$. Astfel, robotul va fi pe deplin convins că propoziţia-π_1, $G(\mathbb{Q}(\mathbf{M}))$, rezultă din ipoteza $\mathcal{M}$, dar şi (presupunând $\mathcal{M}$) că nu este ceva ce poate fi perceput indubitabil fără a-l utiliza pe $\mathcal{M}$ (deoarece nu aparţine lui $\mathbb{Q}(\mathbf{M})$). În felul acesta, $G(\mathbb{Q}(\mathbf{M}))$ este o aserţiune-$\bigstar_{\mathcal{M}}$ dar nu este o aserţiune-$\bigstar$

Să considerăm acum sistemul formal $\mathbb{Q}_{\mathcal{M}}(\mathbf{M})$ construit în acelaşi mod ca şi $\mathbb{Q}(\mathbf{M})$, cu excepţia faptului că aici aserţiunile-$\bigstar_{\mathcal{M}}$ joacă rolul aserţiunilor-$\bigstar$ din construcţia lui $\mathbb{Q}(\mathbf{M})$. În concluzie, $G(\mathbb{Q}(\mathbf{M}))$ este o aserţiune $\bigstar_{\mathcal{M}}$ dar nu este o aserţiune-$\bigstar$ Înseamnă că teoremele din $\mathbb{Q}_{\mathcal{M}}(\mathbf{M})$ sunt sau (i) aserţiunile $\bigstar_{\mathcal{M}}$ însele, sau (ii) propoziţii obţinute din aceste aserţiuni $\bigstar_{\mathcal{M}}$ utilizând logica elementară

[7] Evident, posibilitatea (d) nu este luată aici în considerare, deoarece robotul este de fapt înzestrat cu M şi, deocamdată, îl considerăm pe M lipsit de elemente pur aleatoare, aşa încât nici (c) nu este luat în considerare.

(conform cu paragraful 3.13). În acelaşi mod, în ipoteza $\mathcal{M}$ robotul este fericit că $\mathbb{Q}$(M) încapsulează convingerile sale indubitabile privind adevărul propoziţiilor-π_1, robotul va fi la fel de fericit că sistemul $\mathbb{Q}_{\mathcal{M}}$(M) încapsulează convingerile lui ferme legate de adevărul propoziţiilor-π_1 care sunt condiţionate de ipoteza $\mathcal{M}$.

Şi acum, să presupunem că robotul tocmai contemplă propoziţia Gödel $G(\mathbb{Q}_{\mathcal{M}}(\mathrm{M}))$. Robotul va fi cu siguranţă convins, indubitabil, că această propoziţie-π_1 este o consecinţă a corectitudinii lui $\mathbb{Q}_{\mathcal{M}}$(M). El va crede de asemenea, indubitabil, în corectitudinea lui $\mathbb{Q}_{\mathcal{M}}$(M) consecinţă a lui $\mathcal{M}$ şi, deoarece este fericit că $\mathbb{Q}_{\mathcal{M}}$(M) încapsulează convingerile sale indubitabile despre capacitatea sa de a deriva propoziţii-π_1 pe baza ipotezei $\mathcal{M}$(El ar explica aşa: 'Dacă accept $\mathcal{M}$, atunci accept toate propoziţiile-π_1 care generează sistemul $\mathbb{Q}_{\mathcal{M}}$(M)). Astfel trebuie să accept că $\mathbb{Q}_{\mathcal{M}}$(M)) este corect, pe baza presupunerii $\mathcal{M}$. Ca urmare, trebuie să accept că $G(\mathbb{Q}_{\mathcal{M}}(\mathrm{M}))$ este adevărat, pe această bază lui $\mathcal{M}$.)

Dar crezând (indubitabil) că propoziţia-π_1 Gödel, $\Omega(\mathbb{Q}_{\mathcal{M}}(\mathrm{M}))$, este o consecinţă a lui $\mathcal{M}$, el trebuie să creadă că $G(\mathbb{Q}_{\mathcal{M}}(\mathrm{M}))$ este o teoremă a lui $\mathbb{Q}_{\mathcal{M}}$(M). Poate crede acest lucru numai dacă crede şi că $\mathbb{Q}_{\mathcal{M}}$(M) este incorect — o contradicţie hotărâtă a faptului că acceptă pe $\mathcal{M}$.

S-a presupus, în mod implicit, în raţionamentul de mai sus, că încrederea indubitabilă a robotului este cu adevărat corectă — deşi, ceea ce se cere în realitate este ca robotul să creadă ca sistemul său de convingeri este corect. În orice caz, robotul se presupune că are cel puţin nivelul uman de înţelegere matematică; şi aşa cum am explicat în paragraful 3.4, înţelegerea matematică umană trebuie să fie, în principiu, corectă.

S-ar putea să pară că presupunerea $\mathcal{M}$ şi definiţiile aserţiunilor-$\bigstar_{\mathcal{M}}$ sunt oarecum vagi. Totuşi, trebuie subliniat că o aserţiune care este o propoziţie-π_1 este o afirmaţie matematică bine definită. Se poate imagina că cele mai multe dintre aserţiunile-$\bigstar_{\mathcal{M}}$ pe care le-ar putea face robotul sunt de fapt aserţiuni obişnuite-$\bigstar$, deoarece ar fi neobişnuit ca robotului să i se pară util să invoce ipoteza $\mathcal{M}$. O excepţie ar fi $G(\mathbb{Q}(\mathrm{M}))$ la care ne-am referit mai sus, deoarece, aici $\mathbb{Q}$(M) joacă pentru robot rolul presupusei 'maşini se demonstrat teoreme' din paragraful 3.1 şi 3.3.

Fiind înzestrat cu $\mathcal{M}$ robotul are acces la propria lui 'maşină de demonstrat teoreme' şi deşi s-ar putea să nu fie indubitabil convins de corectitudinea acestei 'maşini' (şi chiar nu este), robotul ar putea foarte bine să se

gândească că ar trebui să fie corectă şi să încerce să deducă consecinţele acestei ipoteze.

În acest fel, robotul nu va fi mai aproape de paradox decât a fost Gödel în stare să-i aducă pe oameni, în concordanţă cu citatul din paragraful 3.1. Totuşi, dacă robotul poate contempla *mecanismul* presupus, **M** şi nu doar sistemul formal $\mathbb{Q}(\mathbf{M})$, raţionamentul poate fi repetat pentru a se trece de la $\mathbb{Q}(\mathbf{M})$ la $\mathbb{Q}_{\mathcal{M}}(\mathbf{M})$ a cărui corectitudine va trebui privită ca o consecinţă a ipotezei $\mathcal{M}$. Şi aceasta ne conduce la o contradicţie (cerută) (vezi, de exemplu, paragraful 3.24 pentru alte discuţii despre sistemul $\mathbb{Q}_{\mathcal{M}}(\mathbf{M})$ şi legătura sa aparentă cu 'gândirea paradoxală').

Concluzia este că nici o fiinţă matematic conştientă — adică, fiinţă capabilă de înţelegere matematică pură — nu poate opera conform unui set de mecanisme, pe care le-ar putea recunoaşte, independent dacă acea fiinţă *ştie* cu adevărat că acele mecanisme ar fi unele care i-ar guverna căile către adevărul matematic indubitabil. (Să reamintim de asemenea că 'adevărul său matematic indubitabil' înseamnă că poate fi stabilit matematic — ceea ce înseamnă prin mijloacele 'demonstraţiei matematice', chiar dacă nu al unei demonstraţii 'formale'.)

Mai precis, suntem conduşi de raţionamentul precedent să concluzionăm că nu există nici o mulţime de mecanisme computaţionale cunoscute de robot, lipsite de ingredienţi aleatori puri, pe care robotul să le accepte, cel puţin ca *posibilitate*, drept bază a sistemului său de convingeri matematice — afară de cazul în care robotul ar fi pregătit să accepte că procedura sugerată de mine pentru construirea sistemului formal $\mathbb{Q}(\mathbf{M})$ din mecanismele **M** ar încapsula totalitatea propoziţiilor-π_1 în care el ar crede fără dubii — şi corespunzător, că sistemul formal $\mathbb{Q}_{\mathcal{M}}(\mathbf{M})$ încapsulează totalitatea propoziţiilor-π_1 despre care crede fără dubii că rezultă din ipotezele $\mathcal{M}$. Mai mult, există situaţia în care ingredienţi aleatori ar putea să fie incluşi în mecanismele **M**, dacă ar fi ca robotul să obţină un sistem potenţial consistent de convingeri matematice.

În următoarele paragrafe, 3.17-3.22, mă voi ocupa de aceste 'portiţe de scăpare' Voi discuta problema încorporării ingredienţilor aleatori în mecanismele **M** (posibilitatea (c)), ca parte a discuţiei mai generale asupra lui (b). În vederea abordării mai atente a problemei (b) trebuie mai întâi să reconsiderăm întreaga problemă a 'convingerilor' robotului, despre care, pe scurt, am mai vorbit la sfârşitul paragrafului 3.12.

3.17. Erorile şi 'înţelegerea' robotului

Întrebarea esenţială pe care trebuie să ne-o punem este dacă robotul este gata să accepte că, dacă a fost construit pe baza unui set de mecanisme **M**, atunci sistemul formal $\mathbb{Q}(\mathbf{M})$ încapsulează sistemul său de convingeri matematice privind propoziţiile-π_1 (şi corespunzător pentru $\mathbb{Q}_{\mathcal{H}}(\mathbf{M})$). Esenţial ar fi ca robotul să fie pregătit să creadă că $\mathbb{Q}(\mathbf{M})$ este *corect* — adică el trebuie să creadă că toate propoziţiile-π_1, care sunt aserţiuni-★ , sunt de fapt *adevărate*. Pentru argumentele pe care le-am formulat, se cere ca *orice* propoziţie-π_1, în care robotul poate să creadă indubitabil, trebuie, de fapt, să fie o teoremă din $\mathbb{Q}(\mathbf{M})$ (astfel încât $\mathbb{Q}(\mathbf{M})$ să poată servi la definirea unei 'maşini de demonstrat teoreme' pentru robot, asemănătoare cu sugestia lui Gödel privitoare la matematicieni, din paragrafele 3.1, 3.3). De fapt, *nu* este esenţial ca $\mathbb{Q}(\mathbf{M})$ să aibă chiar acest rol universal pentru capacităţile potenţiale ale robotului în legătură cu propoziţiile-π_1, dar trebuie să fie suficient de larg pentru a cuprinde utilizarea argumentului Gödel, care i-ar permite să fie aplicat sistemului $\mathbb{Q}(\mathbf{M})$ însuşi (şi corespunzător lui $\mathbb{Q}_{\mathcal{H}}(\mathbf{M})$). Vom vedea în cele ce urmează că acesta este un lucru bine definit — necesitând doar să fie aplicat unui sistem finit de propoziţii-π_1.

Astfel noi — şi robotul — trebuie să facem faţă posibilităţii ca aserţiunile-★ să fie uneori eronate, chiar dacă corectabile, de către robot, în conformitate cu propriile sale criterii. Ideea ar fi că robotul ar putea să se comporte foarte asemănător cu modul în care se comportă un matematician uman. Un matematician, cu siguranţă, poate să ajungă într-o situaţie în care el sau ea să creadă că o anumită propoziţie-π_1 a fost stabilită neîndoielnic ca adevărată (sau poate falsă) — în timp ce, în realitate, există o eroare în raţionament, pe care s-ar putea s-o observe mai târziu. Atunci, raţionamentul anterior ar fi considerat greşit în conformitate cu anumite criterii, ca acelea care fuseseră adoptate înainte, pe când eroarea nu fusese încă depistată — iar o propoziţie-π_1 care părea indubitabil adevărată se poate dovedi falsă (sau viceversa).

Ne putem aştepta ca robotul să se comporte într-un mod asemănător, astfel încât s-ar putea întâmpla ca aserţiunile-★ pe care le formulează să conţină erori, chiar dacă robotul le-a marcat cu semnul ★ Mai târziu, robotul ar putea să-şi corecteze eroarea, dar o eroare a fost făcută, fără îndoială. Cum afectează acest lucru concluziile noastre privind corectitudinea sistemului $\mathbb{Q}(\mathbf{M})$? Evident, acum

$\mathbb{Q}(\mathbf{M})$ nu este întru totul corect şi nici nu este 'perceput' ca întru totul corect de către robot, aşa încât propoziţia lui Gödel $G(\mathbb{Q}(\mathbf{M}))$ nu poate fi crezută. În esenţă, acest lucru este implicat în (b).

Să reconsiderăm acum întrebarea despre ce poate însemna pentru robotul nostru să ajungă la o concluzie matematică 'indubitabilă'. Putem compara situaţia cu ceea ce se întâmplă în cazul unui matematician uman. Într-un astfel de caz, nu ne interesează ce anume s-a întâmplat să afirme matematicianul, practic, ci ce ar trebui luat, în principiu, drept adevăr indubitabil. Să ne amintim de vorbele lui Feynman: „Nu asculta ce spun; încearcă să pricepi ce-am vrut să zic" Pare că ar trebui să ne preocupăm de ceea ce robotul nostru *înţelege*, mai degrabă decât de ceea ce spune. Dar, dacă cineva susţine punctul de vedere $\mathcal{B}$ mai degrabă decât $\mathcal{A}$, nu este clar cum ar putea interpreta, de exemplu, ideea că, pentru robot, ceva ar avea *semnificaţie*. Dacă ar fi posibil să te bazezi nu pe ceea ce robotul afirmă cu ★ ci pe ceea ce, de fapt, 'înseamnă', sau ceea ce, în principiu, ar trebui să însemne — atunci problema unei posibile imprecizii din aserţiunile-★ ar fi circumscrisă. Dificultatea este că nu există nici o metodă de a accesa din exterior astfel de 'semnificaţii' sau intenţii de 'semnificaţii'. În ceea ce priveşte sistemul formal $\mathbb{Q}(\mathbf{M})$, rezultă că trebuie să ne bazăm doar pe aserţiunile-★ propriu zise şi nu putem fi complet siguri de justeţea lor.

Poate fi percepută oare aici o posibilă distincţie operaţională între implicaţiile punctelor de vedere $\mathcal{A}$ şi $\mathcal{B}$? Probabil că da; deoarece, deşi $\mathcal{A}$ şi $\mathcal{B}$ sunt echivalente cu privire la ce anume poate atinge extern un sistem fizic, în principiu, persoanele care susţin aceste puncte de vedere ar putea să se deosebească prin *aşteptările* lor legate de tipul sistemului de calcul pe care-l concep ca fiind capabil de simularea efectivă a creierului unei persoane — exact când percepe validitatea unui argument matematic (vezi şi sfârşitul paragrafului 3.12). Asemenea diferenţe în aşteptări nu par foarte semnificative pentru argumentaţia de faţă.

3.18. Cum poate fi încorporat elementul aleator în activitatea robotului?

În absenţa unei căi efective către aceste chestiuni de semantică, trebuie să ne bazăm pe aserţiunile-★ pe care robotul nostru le face în concordanţă cu

mecanismele care îi comandă comportarea. Trebuie să acceptăm că unele dintre acestea ar putea conține erori, dar astfel de erori ar fi corectabile și, în ori ce caz, ar fi foarte rare. Ar fi rezonabil să presupunem că ori de câte ori un robot face o greșeală în una dintre aserțiunile sale ★, această eroare poate fi atribuită, cel puțin în parte, unor factori întâmplători din mediul exterior sau din activitatea sa internă. Dacă ne imaginăm un al doilea robot, operând după exact aceleași tipuri de mecanisme ca primul, dar pentru care acești factori sunt diferiți, atunci al doilea robot nu va face exact aceleași greșeli ca primul — deși s-ar putea să facă altele. Aceste evenimente întâmplătoare ar putea să fie exact ingredienții aleatori specificați ca parte a intrărilor robotului din mediul înconjurător, sau ca parte a activității interne a robotului. În altă variantă ei ar putea fi pseudo-aleatori, externi sau interni și rezultat al unui calcul determinist dar haotic.

Pentru scopurile argumentației de față, voi presupune că nici un ingredient pseudo-aleator nu are vreun rol de jucat, altul decît acela care ar fi fost jucat, cel puțin tot atât de bine, de către un ingredient pur aleator. Cu siguranță, acesta este punctul de vedere normal. Totuși, mai rămâne o posibilitate, aceea că ar putea exista ceva în comportamentul unui sistem haotic — mergând mai departe de rolul lui de simulare a întâmplării — care să aproximeze o comportare non-computațională. Eu n-am văzut niciodată acest lucru bine analizat, deși multă lume și-a pus speranța în modelul comportării haotice pentru activitatea creierului. Pentru mine personal, astfel de argumente rămân neconvingătoare până când o anumită comportare nealeatoare (adică non-pseudo-aleatoare) a unor astfel de sisteme ar putea fi demonstrată — o comportare care într-un sens tare să aproximeze o comportare non-computațională pură. Până acum, nu am găsit nici o indicație despre o astfel de demonstrație. Mai mult, așa cum voi preciza mai târziu (în paragraful 3.22), ar fi, în ori ce caz, improbabil ca o comportare haotică să poată scăpa de aceste dificultăți pe care argumentele de tip Gödel le ridică pentru modelele computaționale ale minții.

Să considerăm, deocamdată, că orice ingredient pseudo-aleator (sau haotic) din robotul nostru sau din mediul său exterior poate fi înlocuit cu unul pur aleator, fără nici o pierdere a eficacității. Pentru a discuta rolul elementului aleator pur, trebuie să considerăm ansamblul tuturor dificultăților. Deoarece presupunem că robotul nostru este comandat digital și că mediul său exterior poate fi de asemenea tratat ca furnizând date de intrare (să ne reamintim partea 'internă' și 'externă'

a benzii maşinii Turing, aşa cum am descris-o în paragraful 1.8), vor exista un număr finit de astfel de alternative. Deşi acest număr poate fi foarte mare, descrierea lui este numai o problemă de calcul. Astfel, întreg ansamblul de posibili roboţi, fiecare acţionând în conformitate cu mecanismele fixate, va constitui el însuşi un sistem de calcul — deşi, fără îndoială, unul care să nu poată fi reprezentat în practică de nici un calculator care ar putea fi imaginat azi. Totuşi, în ciuda incapacităţii actuale da a realiza o simulare combinată a tuturor roboţilor acţionând conform cu mecanismele **M**, calculul în sine nu este 'incognoscibil'; adică, se poate vedea cum trebuie construit un calculator (teoretic) — sau o maşină Turing — care ar putea realiza simularea, chiar dacă ar fi deocamdată în afara posibilităţilor de realizare în fapt. Acesta este un punct important în discuţia noastră. Un calcul cognoscibil sau un mecanism cognoscibil este ceva ce poate fi specificat de o fiinţă umană; nu este nevoie să rezulte despre un calcul că poate fi realizat de o fiinţă umană , sau chiar de un calculator care ar putea fi practic construit. Situaţia aceasta este foarte asemănătoare cu cea de mai sus, în legătură cu **Q8** şi este consistentă cu terminologia introdusă la începutul paragrafului 3.5.

3.19. Suprimarea aserţiunilor-★ eronate

Să ne întoarcem acum la problema erorilor (corectabile) din aserţiunile-★, pe care robotul le-ar putea face ocazional. Să presupunem că robotul nostru face într-adevăr astfel de erori. Dacă presupunem că un alt robot, sau acelaşi robot, dar într-un alt moment — sau în alte împrejurări — nu ar face probabil aceleaşi greşeli, atunci am putea, în principiu, să identificăm o eroare căutând în interiorul ansamblului de roboţi în funcţiune. Am putea prezenta tabloul simulării comportării tuturor acestor roboţi astfel încât toate acţiunile roboţilor, desfăşurate în timp, să fie considerate simultane (acesta este doar un mod convenabil de a prezenta lucrurile; dar nu presupune că simularea noastră ar funcţiona în realitate ca o acţiune 'paralelă'. Aşa cum am văzut mai devreme, nu există nimic special care să deosebească funcţionarea serială de cea paralelă, în afară de eficienţă; vezi şi paragraful 1.5.). Ideea ar fi ca examinând rezultatele simulării, să fie posibilă, în principiu, eliminarea aserţiunilor-★ eronate, din mulţimea celor corecte, utilizând avantajul că aserţiunile eronate sunt 'corectabile' şi de aceea vor fi judecate ca

erori în marea majoritate a instanţierilor robotului simulat — cel puţin ca o experienţă 'paralelă' a instanţierilor pe care robotul le-a experimentat în timp. Nu cerem ca acesta să fie un procedeu practic, ci doar unul computaţional, unde regulile **M**, care stau la baza întregului calcul, să fie lucruri, în principiu, 'cognoscibile'.

De fapt, cu scopul de a face simularea noastră mai aproape de ceea ce este mai potrivit pentru comunitatea umană a matematicienilor şi, de asemenea, pentru a fi de două ori siguri că toate aserţiunile-★ eronate vor fi suprimate, să considerăm că mediul robotului poate fi împărţit într-o *comunitate* a celorlalţi roboţi şi un mediu rezidual fără roboţi (şi fără oameni) — şi că trebuie să permitem existenţa unor profesori în plus, în mediul rezidual, cel puţin în stadiul iniţial al dezvoltării roboţilor, astfel încât, în particular, semnificaţia strictă a utilizării de către robot a semnului tipografic '★ ' să-i fie clară robotului. Toate comportările posibile ale tuturor roboţilor, împreună cu toate mediile reziduale (relevante) şi intrările datorate oamenilor, variind în raport cu diferite alegeri ale parametrilor aleatori, vor face parte, ca diferite instanţieri, din ansamblul simulat. Din nou, regulile — pe care le notăm tot cu **M** — pot fi luate ca lucruri care pot fi cunoscute, în ciuda complicaţiei detaliilor de calcul care ar interveni dacă simularea ar fi realizată cu adevărat.

Vom lua în considerare (în principiu) toate propoziţiile-π_1 care sunt marcate cu ★ — sau a căror negaţii sunt marcate cu ★ — din toate instanţierile roboţilor (simulaţi computaţional). Vom încerca să separăm toate aserţiunile-★ care sunt lipsite de erori. Putem cere ca orice aserţiune-★ despre o propoziţie-π_1 să fie *ignorată*, afară de cazul când, într-o perioadă de timp T, viitoare sau trecută, numărul r al diferitelor instanţieri ale aserţiunii-★ în ansamblul tuturor simulărilor simultane satisface $r > L + Ns$, unde L şi N sunt numere mari, potrivit alese şi unde s este numărul aserţiunilor-★ din acelaşi interval de timp, care iau o poziţie opusă relativ la propoziţia-π_1 sau, pur şi simplu, afirmă că raţionamentul care stă la baza aserţiunii-★ originare este eronat. Putem cere ca perioada de timp T (care nu trebuie, de fapt, să fie un timp simulat efectiv, ci dat de unităţi de activitate de calcul) precum şi L şi N să poată creşte, pe măsură ce creşte 'complicaţia' propoziţiei-π_1, care trebuie analizată. Atributul 'complicat', relativ la propoziţii-π_1, trebuie precizat în termenii specificării unei maşini Turing, aşa cum a fost ea făcută în paragraful 2.6 (şi în răspunsul la **Q8**). Concret, putem utiliza formulările

explicite date în **MN**, capitolul 2 şi reluat în Anexa A (a acestei cărţi). Astfel vom considera *gradul de complicaţie* al unei propoziţii-π_1, care afirmă că un calcul $T_m(n)$ al unei maşini Turing nu se opreşte, ca fiind numărul ρ al cifrelor binare al celui mai mare dintre numerele n şi m.

Motivul pentru care este inclus aici L, este că trebuie ţinut seama de următoarea posibilitate. Să presupunem că, foarte rar, din mulţimea noastră de alternative, apare un robot 'prost' care face o aserţiune-★ absolut ridicolă, care n-ar putea fi făcută niciodată de vreun alt robot o aserţiune atât de absurdă încât nu a mai apărut niciodată spre a fi judecată de vreunul dintre roboţi! O astfel de aserţiune-★ ar putea fi considerată ca lipsită de erori, în conformitate cu criteriile noastre dacă nu-l includem pe L. Cu un L suficient de mare, o astfel de posibilitate nu ar putea să apară, dacă presupunem că o astfel de 'prostie' la un robot este foarte rară. (S-ar putea ca să fi scăpat alte posibilităţi de acest tip, precum şi alte precauţii de care poate fi nevoie; dar, pentru moment, pare rezonabil să lucrăm pe baza criteriilor pe care le-am sugerat.)

Ţinând cont de faptul că aserţiunile-★ au fost presupuse afirmaţii 'indubitabile' făcute de robot bazate pe un raţionament logic clar, aflat la dispoziţia robotului, din care s-a exclus tot ce ar fi putut fi perceput ca dubios pare rezonabil ca eventualele scăpări în raţionament să fie cu adevărat eliminate în acest mod, unde funcţiile $T(\rho)$, $L(\rho)$ şi $N(\rho)$ nu au nevoie de nimic neobişnuit. Dacă lucrurile stau aşa, suntem din nou în posesia unui sistem *computaţional* un sistem care poate fi cunoscut (în sensul că regulile care stau la baza comportării robotului sunt cognoscibile), presupunând că mecanismele originare **M** care stau la baza comportării robotului pot fi cunoscute. Acest sistem computaţional ne furnizează un sistem formal $\mathcal{Q}'(\mathbf{M})$ ale cărui teoreme sunt aserţiunile-★ , lipsite de erori (sau aserţiuni care pot fi obţinute din acestea prin operaţiile logice ale calculului predicatelor).

În realitate, ceea ce va fi esenţial pentru scopurile noastre nu sunt aceste aserţiuni presupuse fără erori, ci faptul că ele sunt 'crezute', de către roboţii înşişi, ca fiind fără erori (să nu uităm că pentru susţinătorii punctului de vedere $\mathcal{K}$ ideea unui robot care 'să creadă' ceva trebuie luată într-un mod absolut opera-ţional, în sensul *simulării* unei astfel de 'credinţe', ca în paragrafele 3,12, 3.17).

Mai precis, ceea ce se cere este ca roboţii să fie pregătiţi să poată crede, în contextul mecanismelor **M**, cele care stau la baza comportării lor ipoteza $\mathcal{M}$

din paragraful 3.16 — că aceste aserţiuni-★ sunt fără erori. Până în acest punct, în această secţiune, m-am ocupat de posibilitatea îndepărtării eventualelor erori din aserţiunile-★ ale robotului.

Dar ceea ce ne preocupă de fapt, în lumina contradicţiei prezentate în paragraful 3.16, este îndepărtarea erorilor din aserţiunile-★$_{\mathcal{M}}$, acele propoziţii-π_1 despre care robotul crede că sunt fără îndoială deduse din $\mathcal{M}$. Cum acceptarea sistemului $\mathbb{Q}'$(**M**) de către robot este condiţionată, în orice caz, de $\mathcal{M}$, noi putem să permitem contemplarea unui sistem formal mai larg, $\mathbb{Q}'_{\mathcal{M}}$(**M**) definit analog cu sistemul formal $\mathbb{Q}_{\mathcal{M}}$(**M**) din paragraful 3.16. Aici $\mathbb{Q}'_{\mathcal{M}}$(**M**) desemnează sistemul formal construit din aserţiuni-★$_{\mathcal{M}}$, care sunt validate ca 'lipsite de erori', în conformitate cu criteriile *T, L, N*, date anterior. În particular, aserţiunea că $G(\mathbb{Q}'_{\mathcal{M}}(\mathbf{M}))$ ar fi adevărată este validată ca o aserţiune-★, lipsită de erori. Cu acelaşi raţionament, ca cel din paragraful 3.16, vedem că roboţii nu pot accepta că au fost construiţi în conformitate cu **M** (împreună cu limitele de validare *T,L,N*), indiferent ce reguli de calcul **M** li s-ar sugera!

Este acest lucru suficient pentru contradicţia noastră? Cititorul ar putea avea o greutate pe inimă, nefiind sigur cât de atenţi am fost, dacă nu cumva n-ar putea să se afle, vreo eroare în vreo aserţiune-★ sau în una ★$_{\mathcal{M}}$. Pentru argumentaţia de mai sus este necesar ca să fi exclus absolut toate aserţiunile-★$_{\mathcal{M}}$ (sau aserţiunile-★) eronate legate de propoziţiile-π_1. Pentru ca noi (sau roboţii) să fim absolut siguri că $G(\mathbb{Q}'_{\mathcal{M}}(\mathbf{M}))$ este adevărată, se cere corectitudinea sistemului $\mathbb{Q}'_{\mathcal{M}}$(**M**) (cu condiţia $\mathcal{M}$). Această corectitudine cere ca nici o aserţiune-★$_{\mathcal{M}}$ să nu fie inclusă — sau să se creadă că este inclusă în sistem. În ciuda precauţiilor noastre, această cerinţă este improbabilă atât pentru noi, cât şi pentru roboţi — şi numai pentru motivul că numărul posibil al acestor aserţiuni este *infinit*.

3.20. Este nevoie doar de un număr finit de aserţiuni-★$_{\mathcal{M}}$

Această problemă particulară poate fi eliminată şi să ne îndreptăm atenţia spre o mulţime *finită* de aserţiuni-★$_{\mathcal{M}}$. Argumentele sunt oarecum tehnice, dar ideea de bază este că trebuie să considerăm doar un număr finit de propoziţii-π_1 ale căror

specificații sunt, într-un anumit sens, 'scurte' Gradul 'scurtimii' necesare, depinde de cât de complicat este **M**. Cu cât este mai complicată specificarea lui **M**, cu atât mai lungi pot fi propozițiile-π_1. 'Lungimea maximă' este dată în termenii unui anumit număr c, care poate fi determinat din gradul de complicație al regulilor care definesc sistemul formal $\mathbb{Q}'_{\mathcal{M}}(\mathbf{M})$. Ideea este că, atunci când trecem la propoziția Gödel a sistemului formal — pe care de fapt trebuie să-l modificăm puțin — obținem ceva ce nu este mult mai complicat decât este chiar acest sistem modificat însuși. În acest fel, cu un anumit exercițiu în alegerea lui c, putem fi siguri că această propoziție Gödel este ea însăși scurtă. Acest lucru ne permite să atingem contradicția cerută fără a ieși din mulțimea finită a propozițiilor-π_1, scurte.

Vom vedea cum obținem acest lucru în mai multe detalii în restul acestui capitol. Aceia dintre cititori care nu sunt interesați de aceste detalii — și sunt sigur că sunt mulți — îi sfătuiesc să sară peste această parte!

Va trebui să modificăm sistemul formal $\mathbb{Q}'_{\mathcal{M}}(\mathbf{M})$ și să trecem la sistemul puțin diferit $\mathbb{Q}'_{\mathcal{M}}(\mathbf{M},c)$ — pe care îl voi nota, pentru simplitate, cu $\mathbb{Q}(c)$ (aruncând acele adausuri care au devenit prea greu de scris!). Sistemul $\mathbb{Q}(c)$ este definit în următorul fel: singurele aserțiuni-$\bigstar_{\mathcal{M}}$ care vor fi acceptate ca 'lipsite de erori' în construcția lui $\mathbb{Q}(c)$ vor fi acelea al căror grad de complicație, pe care l-am notat cu ρ, este mai mic decât c, unde c este un număr potrivit ales, într-un mod pe care urmează să-l explic. Mă voi referi la acele aserțiuni-$\bigstar_{\mathcal{M}}$ pentru care $\rho < c$, ca la aserțiuni-$\bigstar_{\mathcal{M}}$ 'scurte' Ca și mai înainte, teoremele din $\mathbb{Q}(c)$ nu vor fi chiar aserțiunile-$\bigstar_{\mathcal{M}}$ 'scurte', ci vor cuprinde și aserțiuni care se pot obține din aserțiile-$\bigstar_{\mathcal{M}}$ 'scurte' prin aplicarea operațiilor standard ale logicii (să zicem ale calculului predicatelor). Deși teoremele lui $\mathbb{Q}(c)$ vor fi în număr infinit, ele vor fi generate de un număr finit de aserțiuni-$\bigstar_{\mathcal{M}}$ 'scurte'. Acum, dacă ne restrângem atenția la această mulțime finită, putem să presupunem că funcțiile T, L și N sunt *constante* (să zicem valoarea maximă a lui ρ). În acest fel, sistemul formal $\mathbb{Q}(c)$ va depinde numai de cele patru numere fixate c, T, L și N și de sistemul general de mecanisme **M** care stau la baza comportării robotului.

Punctul esențial al acestei discuții este că procedura lui Gödel este un lucru fix, care cere numai o complicație cu un nivel definit. Propoziția Gödel $G(\mathbb{F})$ pentru sistemul $\mathbb{F}$ este o propoziție-π_1 al cărei grad de complicație cere doar să fie mai mare decât cea a sistemului $\mathbb{F}$ cu o cantitate care poate fi specificată cu precizie.

Pentru a fi mai explicit, în cele ce urmează voi face, oarecum un abuz de notaţie şi voi folosi expresia '$G(\mathbb{F})$' într-un mod special care nu coincide exact cu notaţia din paragraful 2.8. Ne interesează $\mathbb{F}$ numai pentru capacitatea lui de a demonstra propoziţii-π_1. În concordanţă cu această capacitate, $\mathbb{F}$ ne va furniza o procedură algebrică A care este capabilă să stabilească — prin semnalizarea terminării acţiunii lui A — exact acele propoziţii-π_1 care pot fi stabilite utilizând regulile din $\mathbb{F}$. O propoziţie-π_1 este o aserţiune de forma 'acţiunea maşinii Turing $T_p(q)$ nu s-a terminat' — unde putem utiliza codificarea din Anexa A (sau MN, capitolul 2). Ne gândim la A ca acţionând asupra perechii (p,q) ca în paragraful 2.5. Astfel, $A(p,q)$ însuşi se va termina dacă şi numai dacă $\mathbb{F}$ este capabil să stabilească acea propoziţie-π_1 particulară care spune că: '$T_p(q)$ nu se termină'. Procedura din paragraful 2.5 ne va furniza un calcul particular (notat '$C_k(k)$' în paragraful 2.5) care, în ipoteza corectitudinii lui $\mathbb{F}$, ne furnizează o propoziţie-π_1 care depăşeşte capacităţile lui $\mathbb{F}$. *Aceasta* va fi propoziţia-π_1 pe care o voi nota acum cu $G(\mathbb{F})$. Este, în esenţă, echivalentă (pentru un sistem formal $\mathbb{F}$ destul de extins) cu afirmaţia '$\mathbb{F}$ este un sistem consistent', deşi cele două diferă în privinţa detaliilor (vezi paragraful 2.8).

Să presupunem că gradul de complicaţie (aşa cum a fost definit în paragraful 2.6, la sfârşitul răspunsului la **Q8**) a lui A este α, adică numărul cifrelor binare din numărul a, unde $A=T_a$. Atunci, prin construcţia dată explicit în Anexa A, vom găsi că gradul de complicaţie a lui $G(\mathbb{F})$ notat η satisface $\eta<\alpha+210log_2(\alpha+336)$. Pentru nevoile acestei argumentaţii trebuie să definim gradul de complicaţie al sistemului formal $\mathbb{F}$ ca fiind, simplu, cel al lui A, adică α. Cu această definiţie, vedem că gradul de complicaţie suplimentar, apărut la trecerea de la $\mathbb{F}$ la $G(\mathbb{F})$ este mai mic decât cantitatea nesemnificativă computaţional $210log_2(\alpha+336)$.

Ideea este acum să arătăm că, dacă $\mathbb{F} = \mathbb{Q}(c)$, pentru un c suficient de mare, atunci $\eta<c$. Ca atare, va rezulta că propoziţia-π_1, $G(\mathbb{Q}(c))$, trebuie să aparţină lui $\mathbb{Q}(c)$, cu condiţia ca $G(\mathbb{Q}(c))$ să fie acceptată de roboţi cu certitudine-★.

Ne putem asigura că $\gamma<c$ făcând să aibă loc relaţia $c>\gamma+210log_2(\alpha+336)$, unde γ este valoarea lui α când $\mathbb{F}=\mathbb{Q}(c)$. Singura dificultate care ar putea să apară aici este faptul că γ ar depinde de c, chiar dacă nu foarte strâns. Această dependenţă de c apare în două moduri diferite. În primul rând, c oferă limita explicită a gradului de complicaţie al propoziţiilor-π_1 care pot fi calificate ca

aserţiuni-$\bigstar_{\mathcal{M}}$ lipsite de erori' în definirea lui $\mathbb{Q}(c)$; în al doilea rând, prin faptul că sistemul $\mathbb{Q}(c)$ depinde în mod explicit de alegerea numerelor T, L şi N şi, s-ar putea crede că, pentru aserţiunile-$\bigstar_{\mathcal{M}}$ de o complicaţie potenţial mai mare, s-ar cere un criteriu mai strict pentru acceptarea aserţiunilor-$\bigstar_{\mathcal{M}}$ 'lipsite de erori':

Cu privire la această dependenţă de c, să observăm că specificarea actuală a numărului c trebuie să fie dată numai o dată (şi de aceea ne vom referi în interiorul sistemului la ea, cu ' c '). Dacă pentru notaţia lui c se foloseşte notaţia binară, atunci γ va depinde logaritmic de c, pentru valori mari ale lui c (numărul cifrelor binare ale unui număr n fiind aproximativ $log_2 n$). Dar cum în realitate ne preocupăm de c pentru a avea o limită şi nu pentru a-l avea pe c ca un număr precis, putem să încercăm o altă abordare. De exemplu, numărul

$$2^{2^{\cdot^{\cdot^{2}}}}$$

cu un şir de s exponenţi, poate fi notat cu s simboluri şi se pot da exemple unde mărimea numărului de specificat creşte cu s încă mai rapid. Astfel, pentru o limită suficient de mare a lui c sunt necesare doar câteva simboluri pentru a specifica această limită.

Cu privire la dependenţa lui T, L, N de c, pare clar că în virtutea consideraţiilor de mai sus, putem din nou să ne asigurăm că specificarea valorilor acestor numere (ca limite extreme) nu cere un număr de cifre binare care să crească rapid cu c şi o dependenţa logaritmică de c ar fi suficientă. Deci putem, evident, presupune că dependenţa lui $\gamma+210log_2(\gamma+336)$ de c nu este mai mult decât logaritmică şi se poate alege astfel încât c să fie mai mare decât acest număr.

Să alegem un astfel de c; şi să notăm acum pe $\mathbb{Q}(c)$, simplu, cu $\mathbb{Q}^*$ Astfel, $\mathbb{Q}^*$ este un sistem formal ale cărui teoreme sunt exact acele propoziţii matematice care pot fi obţinute, utilizând regulile standard ale logicii (calculului predicatelor), din aserţiunile $\bigstar_{\mathcal{M}}$'scurte', în număr finit. Aceste aserţiuni-$\bigstar_{\mathcal{M}}$ sunt în număr finit, astfel încât o mulţime de numere fixe T, L, N vor fi suficiente pentru ca să garanteze că ele sunt lipsite de erori. Dacă roboţii cred acest lucru, cu o certitudine-$\bigstar_{\mathcal{M}}$, atunci ei ar trebui să concluzioneze-$\bigstar_{\mathcal{M}}$ că propoziţia Gödel $G(\mathbb{Q}^*)$ este de asemenea adevărată pe baza ipotezei $\mathcal{M}$, aceasta fiind o propoziţie-π_1 cu o complicaţie mai mică de c. Argumentaţia care deduce $G(\mathbb{Q}^*)$ din convingerea-$\bigstar_{\mathcal{M}}$ în corectitudinea sistemului $\mathbb{Q}^*$ este una simplă (în principal cea

pe care tocmai am dat-o) așa încât nu vor exista probleme cu validarea-$\bigstar_{\mathcal{M}}$ sa. Astfel, $G(\mathbb{Q}^*)$ va trebui să fie ea însăși o teoremă în $\mathbb{Q}^*$ Dar acest lucru contrazice credința roboților în corectitudinea lui $\mathbb{Q}^*$. Astfel, această convingere (presupunând $\mathcal{M}$ și faptul că numerele T, L, N sunt suficient de mari) ne conduce la o inconsistență cu mecanismele **M** care stau de fapt la baza funcționării robotului — cu implicația că **M** *nu poate* sta la baza funcționării robotului.

Dar cum ar putea fi roboții siguri că numerele T, L, N au fost alese suficient de mari? Ei ar putea să nu fie siguri, dar atunci ceea ce ar putea face ar fi să aleagă un set de valori T, L și N și să încerce să presupună că sunt suficiente — de unde vor obține o contradicție cu ipoteza de pornire cum că ei acționează în conformitate cu mecanismele **M**. Apoi ar putea să presupună că niște valori cu ceva mai mari ar putea fi suficiente — caz în care ar ajunge din nou la o contradicție — și așa mai departe. Ar realiza curând că se obține o contradicție pentru *orice* valori alese (cu observația că pentru valori absolut stupefiante pentru T, L și N, valoarea lui c ar putea și ea să crească puțin — dar acesta nu este un lucru important).

Astfel, se obține aceeași concluzie indiferent de valorile T, L și N, așa încât robotul ajunge la concluzia — așa cum aparent am ajuns și noi — că nici un fel de procedură computațională, cognoscibilă **M** nu poate sta la baza procesului de gândire matematică!

3.21. Sunt suficiente măsurile de prevedere luate?

Să observăm că această concluzie rezultă pentru o clasă foarte largă de posibile sugestii pentru garanți ai corectitudinii. Nu este necesar ca aceștia să aibă forma pe care am propus-o aici. Ne putem imagina că unele îmbunătățiri ar putea fi necesare. De exemplu, probabil că există o tendință a roboților să devină 'senili' după ce au rulat o lungă perioadă de timp și comunitatea lor are tendința de a se degrada și standardele au tendința să cadă, așa încât mărind numărul T dincolo de un punct, să creștem de fapt șansele de eroare în aserțiunile $\bigstar_{\mathcal{M}}$!

O altă situație poate să apară dacă N (sau L) sunt luați prea mari, atunci s-ar putea cere rularea tuturor aserțiunilor-$\bigstar_{\mathcal{M}}$ pe care le fac niște roboți 'nebuni', din când în când și care nu sunt întrecute, ca număr, destul de categoric, de

aserţiunile-★ făcute de roboţii cu bun simţ. Fără îndoială, nu ar fi greu să eliminăm acest fel de lucruri punând alte limitări pentru parametri sau, să zicem, având o elită a roboţilor, unde roboţii membri să fie continuu testaţi pentru a fi siguri de capacităţile lor mentale — şi insistând ca semnul tipografic ★ să fie pus numai după aprobarea societăţii întregi.

Există multe alte posibilităţi de a îmbunătăţi calitatea aserţiunilor-$\bigstar_m$ pentru a le elimina pe cele eronate dintre toate celelalte (în număr finit). Unii s-ar putea îngrozi că, deşi limita c a complicaţiei propoziţiilor-π_1 ne conduce la un număr finit de candidaţi pentru starea-★ sau starea-$\bigstar_m$, numărul este încă enorm de mare (depinzând exponenţial de c), aşa încât este greu să fim siguri că toate aserţiunile-$\bigstar_m$ care ar putea fi eronate au fost îndepărtate. Într-adevăr, nu s-a specificat nici o limită pentru numărul de paşi de calcul care ar fi necesari robotului pentru a oferi o demonstraţie-$\bigstar_m$ satisfăcătoare pentru o astfel de propoziţie-π_1. Trebuie să fie clar că, cu cât este mai lung lanţul raţionamentului dintr-o astfel de demonstraţie, cu atât mai stringentă este nevoia unui criteriu care să stabilească starea-$\bigstar_m$ pentru respectiva demonstraţie. Acesta este, în cele din urmă şi modul în care reacţionează, în general, matematicienii. O demonstraţie lungă şi complicată va necesita o mare atenţie şi multă grijă, înainte de a putea fi acceptată ca o demonstraţie indubitabilă. La fel s-ar întâmpla în cazul în care o demonstraţie ar fi examinată de roboţi, pentru a fi considerată, eventual, ca o stare-$\bigstar_m$.

Argumentele aduse mai sus vor rămâne valabile pentru orice fel de modificări ale propunerilor de aici de excludere a erorilor, cu condiţia ca natura acestor modificări să fie oarecum similară cu a celor pe care le-am sugerat aici. Tot ceea ce este necesar pentru ca argumentaţia să meargă, este ca astfel de propuneri clar formulate care să excludă toate aserţiunile-$\bigstar_m$ eronate, să existe. Ajungem la concluzia riguroasă:

nici un fel de mecanism computaţional, garantat şi cognoscibil nu poate încapsula raţionamentul matematic uman.

Ne-am preocupat de aserţiuni-$\bigstar_m$ care, deşi uneori, ocazional, ar putea fi eronate, sunt în principiu *corectabile* de către roboţi — chiar dacă nu sunt de fapt corectate în anumite situaţii din existenţa simulată a roboţilor. Este greu de văzut ce înseamnă erori 'în principiu corectabile' (operaţional), dacă nu corectabile în raport cu o procedură generală ca cea expusă mai sus. O eroare care nu a fost corectată de robotul care a făcut-o poate fi corectată de un alt robot — mai mult,

în cele mai multe ocazii ale existenței potențiale a roboților, acea eroare particulară nu va mai fi făcută niciodată.

Concluzia (cu unele schimbări minore pe care le-ar cere considerarea ingredienților haotici, în locul celor aleatori; ca în paragraful 3.22) este că nici un set de reguli de calcul **M**, de natură *top-dowm*, *bottom-up* sau *autocorectoare* sau vreo combinație a lor, nu poate sta la baza comportării comunității noastre de roboți sau a vreunui membru, robot individual — dacă ar fi să presupunem că ei pot atinge nivelul înțelegerii matematice umane!

Dacă ne imaginăm pe noi înșine acționând ca niște roboți comandați computațional suntem efectiv conduși la o contradicție.

3.22. Poate haosul să salveze modelul computațional al minții?

Mă voi întoarce acum, pe scurt, la problema haosului. Deși, așa cum a fost deja subliniat de mai multe ori în această carte (în particular în paragraful 1.7), sistemele haotice sunt simple cazuri particulare de sisteme de calcul; există o concepție conform căreia fenomenul haosului ar putea avea relevanță pentru funcționarea creierului. În discuția de mai sus, am considerat, la un anumit moment, părerea aparent rezonabilă că orice comportare computațională haotică ar putea fi înlocuită cu una pur aleatoare, fără pierderi semnificative. Ar trebui să reexaminăm această presupunere. Comportarea unui sistem haotic — deși în mod normal poate implica complicații mari pentru a părea aleatoare — nu este de fapt aleatoare. Într-adevăr, unele sisteme haotice se comportă în moduri foarte ciudate și complexe, lucru care se datorează, în mod special, purei întâmplări (uneori expresia 'la marginea haosului' este utilizată pentru a descrie comportarea nealeatoare, complicată care ar putea să apară în sistemele haotice [10]). Ar putea, ceea ce se cheamă haos, să ofere răspunsul dorit pentru misterul activității mentale? Pentru aceasta ar trebui să apară o înțelegere complet nouă a modului în care se comportă sistemele haotice în anumite situații. S-ar putea întâmpla ca, în astfel de situații, un sistem haotic să poată aproxima bine o *comportare non-computațională* la limită — sau ceva de această natură. Nici o astfel de demonstrație nu a fost, după cunoștințele mele, oferită până acum. Totuși ea rămâne o posibilitate

interesantă şi eu sper că, în anii ce vin, ea va fi realizată.

Independent de această posibilitate, haosul ar putea oferi doar o foarte slabă şansă de a evita concluzia la care am ajuns în secţiunea precedentă. Haosul nealeator (adică non-pseudoaleator) a jucat, totuşi, un rol în discuţia anterioară, permiţându-ne să considerăm simularea nu numai a comportării robotului nostru (sau a comunităţii de roboţi) ci şi simularea întregului ansamblu de posibile activităţi ale robotului, consistentă cu mecanismele **M** date.

Elemente aleatoare ar putea fi implicate în datele iniţiale care oferă punctul de pornire pentru simulare şi mai putem utiliza aceeaşi idee pentru a manevra aspectul aleator în furnizarea numărului mare de posibilităţi alternative pentru istoriile robotului, în simularea simultană. Dar comportarea haotică în sine, va trebui pur şi simplu *calculată* — aşa cum comportarea haotică este în mod normal calculată în practică, pe un calculator, în cazul exemplelor numerice. Ansamblul alternativelor posibile nu va fi chiar atât de mare încât să fie legitimă aproximarea haosului prin aleatoriu. Dar singurul motiv pentru care am considerat astfel de ansambluri mari a fost pentru a fi de două ori asiguraţi că posibilele greşeli din aserţiunile-$\bigstar_{\mathcal{M}}$ ale robotului sunt cu siguranţă eliminate. Chiar dacă ansamblul considerat este doar istoria unui singur robot din comunitate, putem fi siguri că, cu un set de criterii destul de restrictive, pentru acceptarea-$\bigstar_{\mathcal{M}}$, astfel de erori ar fi îndepărtate de către ceilalţi roboţi din comunitate, sau de robotul însuşi, cândva, mai târziu. Cu un ansamblu suficient de mare, format din elemente pur aleatoare, eliminarea erorilor ar fi mai efectivă, dar lărgirea mulţimii printr-o aproximare aleatoare care să înlocuiască comportarea pur haotică nu pare prea importantă. În concluzie, haosul nu ne scoate din dificultatea în care am ajuns cu modelul computaţional.

3.23. *Reductio ad absurdum* — un dialog fantastic

Multe dintre argumentele din paragrafele precedente din acest capitol au fost oarecum complicate. Cu scopul de a face o prezentare rezumativă voi prezenta o conversaţie imaginară, care are loc peste mulţi ani, în viitor, între un presupus specialist practician IA de mare succes şi robotul cu multe premii, creat de el. Povestea este scrisă din punctul de vedere IA tare.

Observaţie: în naraţiune, Q joacă rolul algoritmului A, utilizat în argumentaţia din paragraful 2.5 şi $G(Q)$ are rolul calculului $C_k(k)$ care nu se opreşte. Raţionamentul acestui paragraf poate fi urmărit având ca bază cel puţin paragraful 2.5.

Împăratul Albert avea toate motivele să fie mulţumit de rezultatele muncii sale de o viaţă. Procedurile pe care le pusese la punct cu mulţi ani înainte dăduseră în sfârşit roade. Şi acum tocmai este angajat într-o discuţie cu una dintre cele mai impresionante creaturi ale sale: un robot cu capacităţi matematice extraordinare şi potenţial supraumane care se cheamă Cybersistemul care Face Matematică (figura 3.2). Antrenamentul robotului era aproape terminat.

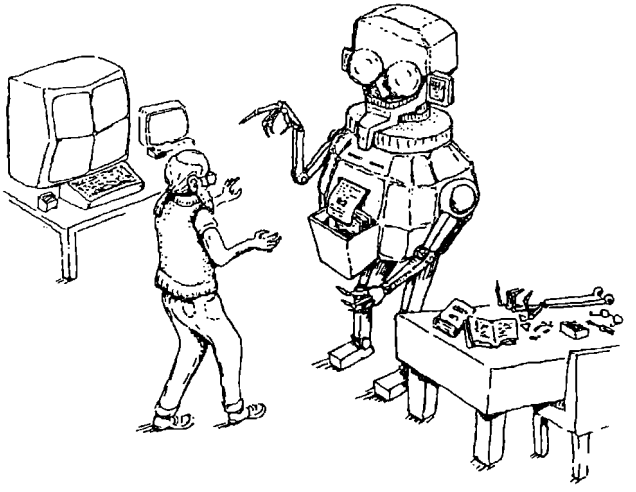


Fig. 3.2. Împăratul Albert înfruntă Cybersistemul de Făcut Matematică

Împăratul Albert: Te-ai uitat peste articolele pe care ţi le-am împrumutat – unele despre Gödel şi alţii, care discută despre implicaţiile acestei teoreme?

Cybersistemul care Face Matematică: Evident deşi articolele erau destul de

elementare, au fost totuşi interesante. Gödel al dumneavostră pare să fi fost un logician destul de capabil — pentru un om.

IA: *Destul* de capabil, doar atât? Gödel a fost cu siguranţă unul dintre cei mai mari logicieni din toate timpurile. Probabil *cel* mai mare!

CFM: Scuzele mele, dacă am lăsat impresia că l-am subestimat. Bineînţeles, aşa cum bine ştiţi, am fost antrenat să fiu în general respectuos cu realizările oamenilor — deoarece oamenii se ofensează uşor — chiar dacă, de obicei aceste realizări ni se par banale, dar am crezut că, cel puţin cu *dumneavoastră,* mă pot exprima direct.

IA: Bineînţeles că poţi. Scuze; m-am pierdut cu firea. Deci n-ai avut nici o dificultate să înţelegi teorema lui Gödel?

CFM: Nici una. Sunt sigur că teorema mi-ar fi venit în minte şi mie dacă aş fi avut ceva mai mult timp. Dar am fost ocupat cu alte probleme mai fascinante despre cohomologia neliniară transfinită, care tocmai mă interesează acum. Teorema lui Gödel mi se pare de bun simţ şi absolut directă. Cu siguranţă, n-am avut nici o dificultate să o urmăresc.

IA: Ha! Asta o încasează Penrose!

CFM: Penrose? Care Penrose?

IA: O, tocmai mă uitam prin carte asta veche. Nu este ceva cu care să fi vrut să te mai încurc. Autorul se pare că a făcut o afirmaţie, cu ceva vreme în urmă, că ceea ce faci tu este un lucru imposibil.

CFM: Ha, ha, ha! *(Robotul face o impresionantă simulare a unui râs batjocoritor)*

IA: Uitându-mă la cartea asta mi-am amintit de ceva. Ţi-am arătat vreodată regulile detaliate pe care le-am utilizat pentru a face procedurile de calcul care au dus la construirea şi dezvoltarea ta şi a celorlaţi colegi ai tăi, roboţi?

CFM: Nu, nu încă. Am sperat că ai putea să o faci odată. Chiar mă întrebam dacă consideri detaliile acestor proceduri ca pe un fel de inutil secret de fabricaţie, sau eşti pur şi simplu jenat de forma lor care în detalii probabil e nefinisată şi ineficientă.

IA: Nu, nu; nu-i deloc aşa. De mult nu mă mai preocup de astfel de chestii. Toate sunt aici în fişiere, pe discurile calculatorului.

După 13 *minute,* 41,7 *secunde mai târziu*

CFM: Fascinant — doar la prima vedere am văzut cel puţin 919 moduri clare în care puteai obţine aceleaşi efecte, mai simplu.

IA: Am fost conştient că ar fi fost posibile unele simplificări, dar ar fi fost mai multă bătaie de cap decât ar fi meritat, încercând să găsim schema cea mai simplă la vremea aceea. Nu ni s-a părut că merita.

CFM: Poate că e adevărat. Nu mă simt prea ofensat că nu ţi-ai dat mai multă silinţă să găseşti schema cea mai simplă. Sper ca nici colegii mei roboţi să nu fie prea ofensaţi.

IA: De fapt cred că am făcut o treabă foarte bună — capacitatea matematică a ta şi a celorlalţi pare cu adevărat impresionantă şi se tot perfecţionează, după câte îmi dau seama. Mai că-mi vine să cred că eşti pe cale să depăşeşti capacitatea matematicienilor oameni.

CFM: E clar că ai dreptate. Pe când vorbeai, tocmai mă gândeam la câteva teoreme care par să fie mult peste rezultatele care se publică în literatura de specialitate a oamenilor. Ba chiar am observat, cu colegii mei, nişte erori în unele dintre rezultatele care erau acceptate ca adevărate de către matematicieni de mulţi ani. În ciuda evidenţei că voi oamenii sunteţi foarte exigenţi cu rezultatele voastre, mă tem că greşelile mai apar din timp în timp.

IA: Dar cu roboţii cum e? Nu credeţi că uneori tu şi colegii tăi aţi putea face

câte o greşeală — înţeleg în teoremele matematice despre care susţineţi că le-aţi stabilit definitiv?

CFM: Nu, cu siguranţă nu. Odată ce un robot matematic a stabilit că un anumit rezultat este o teoremă, se poate considera că rezultatul este fără îndoială adevărat. Noi nu facem genul acela de greşli stupide pe care oamenii le fac ocazional în afirmaţiile lor matematice ferme. Evident, în cercetarea preliminară — ca şi oamenii — adeseori încercăm diverse lucruri şi încercăm să ghicim. Astfel de lucruri se pot dovedi greşite; dar atunci când afirmăm despre ceva că este stabilit matematic, atunci *garantăm* validitatea lui.

După cum ştii, eu şi colegii mei am început să publicăm câteva dintre rezultatele noastre în câteva dintre cele mai prestigioase reviste electronice, dar nu ne-am simţit prea bine să vedem modul neglijent în care colegii tăi matematicienii îşi redactează lucrările. Ne-am propus să începem să publicăm propria noastră 'revistă' — o bază de date cuprinzătoare de teoreme matematice pe care *noi* le acceptăm ca fiind de necontestat demonstrate. Aceste rezultate vor fi marcate cu un semn tipografic ✯ (un simbol pe care parcă chiar tu ni l-ai sugerat mai de mult) şi care să însemne: acceptat de Societatea pentru Inteligenţa Matematică în Comunitatea Roboţilor (SIMCR) — o societate cu criterii foarte riguroase de acceptare şi cu testări continue care să ne asigure că nu au intervenit nici un fel de deteriorări mentale la vreun robot, oricât de neplauzibilă ar putea să-ţi pară ţie (sau nouă) o astfel de posibilitate. Spre deosebire de standardele inferioare pe care oamenii le-au adoptat, poţi fi sigur că atunci când punem semnul nostru ★ lângă un rezultat, noi *garantăm* că este matematic adevărat.

IA: Chiar acum mi-am amintit de ceva din cartea veche de care tocmai ţi-am spus mai înainte. Mecanismele originare $\mathbf{M}_0$ în conformitate cu care eu şi colegii mei am pus pe picioare ceea ce acum formează comunitatea roboţilor matematici — şi nu uita că asta include simularea computaţională a tuturor factorilor de mediu, antrenamentul şi un proces riguros de selecţie la care v-am supus şi un proces *buttom-up* de învăţare prin care aţi trecut — te-ai gândit vreodată că toate astea v-au furnizat *o procedură de calcul* pentru generarea tuturor aserţiunilor pe care le-ar accepta vreodată SIMCR? Este o procedură computaţională pentru că voi, roboţii, sunteţi entităţi computaţionale care aţi evoluat pe baza unui proces de 'selecţie

naturală' care a funcţionat într-un mediu complet computaţional — în sensul că o simulare computaţională a întregii operaţii este posibilă, în principiu. Dezvoltarea întregii voastre societăţi de roboţi reprezintă, îndeplinirea unui calcul extrem de elaborat, iar familia tuturor aserţiunilor-★ cu care vreodată v-aţi putea întâlni este ceva ce poate fi generat de o maşină Turing particulară. Este chiar o maşină Turing pe care aş putea-o, în principiu scrie; de fapt, cred că dacă aş avea câteva luni, aş putea-o face chiar practic, utilizând fişierele pe care tocmai ţi le-am arătat.

CFM: Pare o observaţie complet elementară. Sigur că ai putea-o face, în principiu, ba chiar cred ca ai putea-o face şi practic, dar e păcat să pierzi atâta vreme din timpul care pentru tine este atât de preţios; eu aş face-o imediat dacă ai vrea să mi-o ceri.

IA: Nu, nu, n-are nici un rost. Dar chiar acum aş vrea să ne mai gândim la lucrurile astea. Să ne concentrăm asupra aserţiunilor-★ care sunt propoziţii-π_1. Îţi aduci aminte ce sunt propoziţiile-π_1?

CFM: Bineînţeles, sunt conştient ce anume este o propoziţie-π_1. Este o afirmaţie că o acţiune a unei anumite maşini Turing nu se opreşte.

IA: OK! Să ne referim acum la o procedură computaţională care generează propoziţii-π_1 asertate-★, ca la Q(M) sau numai Q, pe scurt. Rezultă că trebuie să existe o aserţiune matematică de tip Gödel — altă propoziţie-π_1, pe care am s-o numesc[8] G(Q) — iar adevărul lui G(Q) este o consecinţă a aserţiunii că roboţii nu fac niciodată greşeli în legătură cu propoziţiile-π_1 pe care eşti gata să le afirmi cu certitudine-★

CFM: Da, s-ar putea să ai dreptate în această hmm.

IA: Şi G(Q) trebuie să fie de fapt adevărată, pentru că voi, roboţii, nu faceţi niciodată greşeli când e vorba de aserţiunile-★, cum le notaţi voi.

---

[8] Strict, notaţia $G(\ )$ ' a fost rezervată pentru sisteme formale, în paragraful 2.8, nu pentru algoritmi, dar o să-i permit unui specialist în IA să aibă aici puţină libertate de alegere.

CFM: Evident.

IA: Stai o clipă dar ar rezulta de asemenea că $G(\mathbf{Q})$ ar trebui să fie ceva pe care, voi roboţii să nu-l puteţi percepe ca adevărat — sau cel puţin, nu cu o certitudine-★

CFM: Faptul că noi roboţii am fost de la început construiti respectând **M**, împreună cu faptul că aserţiunile-★ pe care le facem asupra propoziţiilor-π_1 nu sunt niciodată greşite, înseamnă cu siguranţă că propoziţia $G(\mathbf{Q})$ trebuie să fie adevărată. Îmi închipui că te gândeşti deja că ar trebui să-i sfătuiesc pe cei din SIMCR să atribuie lui $G(\mathbf{Q})$ simbolul tipografic-★ , dacă ei oricum acceptă că nu poate interveni nici o eroare la atribuirea simbolului ★ Într-adevăr, ar trebui să accepte asta. Singura poantă cu simbolul ★ este că el oferă o *garanţie* de corectitudine.

Deşi este imposibil să accepte pe $G(\mathbf{Q})$ pentru că, prin chiar natura construcţiei lui Gödel, $G(\mathbf{Q})$ este situat în afara a ceea ce noi putem aserta-★ — cu condiţia ca să nu facem de fapt niciodată greşeli în aserţiunile-★ Cred că şi ţie ţi s-ar putea întâmpla să crezi că ceva aici este neclar în mintea noastră legat de siguranţa asignărilor cu ★ pe care noi le facem.

Totuşi, nu pot concepe că aserţiunile-★ făcute de noi ar putea fi vreodată greşite, cu toate precauţiile şi grja cu care SIMRC lucrează. Cred că numai voi, oamenii, aţi putut greşi şi procedurile încorporate în **Q** nu sunt de fapt cele pe care le folosiţi, chiar în ciuda a ceea ce mi-ai spus şi a ceea ce demonstraţia voastră pare să arate. În ori ce caz, SIMRC nu va fi niciodată sigur de faptul că noi suntem construiţi în conformitate cu **M**, adică prin procedurile încapsulate de **Q**. N-avem, pentru asta decât cuvântul vostru.

IA. Pot să te asigur că procedurile sunt chiar acestea; ştiu, pentru că eram acolo şi răspundeam personal de ele.

CFM: N-aş vrea să crezi că-ţi pun la îndoială cuvântul. Poate că unul dintre asistenţii tăi a greşit pe când îţi executa instrucţiunile. Tipul ăsta, Fred Carruthers — face tot timpul greşeli stupide. Nu m-aş mira dacă ar fi băgat niste erori serioase.

IA: Te îmbeţi cu apă rece! Chiar dacă ar fi introdus vreo eroare, colaboratorii mei şi cu mine i-am fi dat de urmă şi am fi aflat exact ce face Q de fapt. Ceea ce cred că te nelinişteşte, de fapt, este că noi *ştim* — sau cel puţin putem afla — ce proceduri au fost folosite pentru realizarea construcţiei voastre. Acest lucru înseamnă că noi putem, cu un anumit efort să scriem propoziţia $G(Q)$ şi să fim siguri că ea este adevărată — cu condiţia ca voi să nu faceţi cu adevărat nici o greşeală în aserţiunile-★ , pe care le faceţi. Totuşi, *voi* nu puteţi fi siguri că $G(Q)$ este adevărată; sau, în orice caz, nu-i puteţi da certitudinea de care este nevoie pentru a stisface criteriile SIMRC într-o măsură suficientă pentru a-i putea da un statut-★ Acest lucru pare să ne dea nouă, oamenilor, un ultim avantaj asupra voastră, a roboţilor, în principiu, nu în practică, pentru că există propoziţii-π_1 care nouă ne sunt accesibile şi care vouă nu vă sunt. Nu ştiu dacă voi, roboţii, veţi putea face faţă acestei posibilităţi — da, evident, *de asta* ne acuzi atât de vehement că am fi făcut greşeli.

CFM: Nu ne mai atribui nouă motivele voastre, meschine, omeneşti. Este evident că *nu pot accepta* că ar exista astfel de propoziţii-π_1 care ar fi accesibile oamenilor şi nu ne-ar fi inaccesibile nouă, roboţilor. Roboţii matematicieni nu sunt prin nimic inferiori oamenilor care fac matematică — deşi trebuie să accept că orice propoziţie-π_1 care ne este nouă accesibilă vă este accesibilă şi vouă oamenilor, dar cu câtă trudă! Dar nu pot accepta, totuşi, că ar exista propoziţii-π_1 care, în principiu să ne fie nouă *in*accesibile, deşi oamenilor le-ar fi accesibile.

IA: Eu cred că însuşi Gödel a văzut posibilitatea existenţei unei proceduri de calcul, aşa cum este Q, aplicată acum *oamenilor* care fac matematică — procedură pe care el o numea 'maşină de demonstrat teoreme' — care ar putea genera exact propoziţii-π_1 a căror adevăr să fie, în principiu, accesibil oamenilor care fac matematică. Deşi nu mă gândesc că el chiar credea că o astfel de maşină ar fi posibilă *în realitate*, el ne putând să o definească matematic. Ceea ce pare că avem aici este o astfel de 'maşină' care vă este aplicabilă vouă, roboţilor, adică Q, care generează toate propoziţiile-π_1 accesibile roboţilor şi a cărui corectitudine vă este inaccesibilă. În plus, cunoscând procedurile algoritmice care stau la baza construcţiei voastre, noi *înşine* avem acces la acest Q şi putem percepe adevărul său — *cu condiţia* ca să putem fi convinşi de faptul că nu faceţi greşeli în

aserţiunile-$\bigstar$

CFM: {*după o pauză destul de lungă*} OK. Cred că voi *aţi putea* crede că, uneori, membrii SIMRC ar putea face vreo greşeală când atribuie eticheta $\bigstar$ Mai cred şi că SIMRC ar putea să nu fie pe deplin convinşi că atribuie eticheta $\bigstar$ fără nici o eroare. În acest fel, $G(\mathbf{Q})$ ar putea să nu obţină asertarea-$\bigstar$ şi contradicţia ar putea fi ocolită. Observă că acest lucru nu înseamnă că eu aş admite că noi, roboţii, am face vreodată erori în aserţiunile-$\bigstar$ Ci doar că nu putem fi absolut siguri că aşa ceva nu s-ar putea întâmpla.

IA: Vrei să-mi spui că, deşi adevărul este absolut garantat pentru fiecare propoziţie-π_1 individuală, asertată-$\bigstar$, nu există nici o garanţie că n-ar exista vreo eroare la întreaga colecţie? Asta mi se pare în contradicţie cu tot ceea ce ar putea să însemne 'indubitabil adevărat'.

Stai o clipă ar putea asta să aibă ceva în comun cu faptul că există o *infinitate* de propoziţii-π_1? Asta îmi aminteşte puţin de ceea ce se numeşte ω-consistenţă care are, dacă îmi aduc bine aminte, ceva cu propoziţia lui Gödel $G(\mathbf{Q})$.

CFM: {*după o pauză ceva mai lungă*} Nu, nu este deloc aşa. Nu are nimic de-a face cu faptul că numărul propoziţiilor-π_1 posibile ar fi infinit. Ne-am putea gândi numai la propoziţii care să fie 'scurte' într-un anumit sens — în sensul că specificarea maşinii Turing corespunzătoare ar putea fi făcută într-un număr dat, mai mic decât c cifre binare. Nu te mai complic cu detaliile pe care tocmai le-am parcurs. Ceea ce rezultă este că există o mărime fixă a lui c, la care ne putem restrânge atenţia, care depinde de gradul de complicaţie al regulilor lui $\mathbf{Q}$. Cum procedura lui Gödel — care este obţinută din $\mathbf{Q}$ — este un lucru simplu şi clar, nu avem nevoie de o complicaţie mai mare în propoziţiile-π_1 decât cea care este deja conţinută în $\mathbf{Q}$. Astfel, restrângând complicaţia acestor propoziţii la c, acest lucru nu împiedică aplicarea procedurii Gödel. Propoziţiile-π_1, restrânse în acest mod, furnizează o familie finită, deşi foarte mare. Dacă ne restrângem atenţia la astfel de propoziţii-π_1 'scurte', vom obţine o procedură computaţională $\mathbf{Q}^*$ — în esenţă de aceeaşi complicaţie ca şi $\mathbf{Q}$ — care generează propoziţiile-π_1, scurte şi asertate-$\bigstar$ Discuţia se aplică exact ca mai înainte. Fiind dat $\mathbf{Q}^*$, vom putea găsi o altă propoziţie scurtă $G(\mathbf{Q}^*)$, care trebuie să fie adevărată ştiind că toate propoziţiile-π_1

scurte, asertate-★ sunt cu siguranţă adevărate, dar care nu poate, în acest caz, fi ea însăşi, asertată-★ — toate acestea însemnând că tu ai dreptate în afirmaţia ta că mecanismele M sunt cu adevărat acelea pe care le-aţi folosit, un 'fapt' de care eu nu sunt prea convins, dacă pot să spun aşa.

IA: Astfel pare să fi ajuns la acelasi paradox, ca mai înainte, dar acum într-o formă mai tare. Există acum o listă *finită* de propoziţii-π_1, fiecare în parte garantată, dar voi — sau SIMRC, sau cum s-o fi numind — nu sunteţi gata să daţi o garanţie absolută că lista, ca un întreg, nu conţine nici o eroare. Deoarece voi nu garantaţi pentru $G(\mathbf{Q}^*)$, a cărui adevăr este o consecinţă a tuturor propoziţiilor-π_1 din lista considerată adevărată. Cu siguranţă ceva este ilogic, nu ţi se pare?

CFM: Nu pot accepta că roboţii sunt ilogici. Propoziţia $G(\mathbf{Q}^*)$ este numai o consecinţă a altor propoziţii-π_1, dacă am fost cu adevărat construiţi în conformitate cu M. Nu putem însă garanta $G(\mathbf{Q}^*)$ deoarece nu putem garanta că am fost construiţi în conformitate cu M. Nu am decât asigurarea ta verbală în legătură cu modul în care am fost construiţi. Roboţii cu siguranţă nu pot depinde de împrejurarea că oamenii sunt supuşi greşelii.

IA: Îţi mai repet că *aşa* aţi fost construiţi — deşi înţeleg că voi roboţii nu aveţi nici un mod de a stabili cu certitudine că acesta este adevărul. Tocmai faptul că noi *ştim*, ne permite să credem în adevărul propoziţiei-π_1, $G(\mathbf{Q}^*)$, dar, în cazul nostru, apare o altă incertitudine în legătură cu faptul că noi nu suntem chiar atât de siguri, aşa cum păreţi voi, că toate aserţiunile-★ sunt lipsite de erori.

CFM: Te asigur că toate sunt fără erori. Şi suntem foarte siguri de asta. Standardele noastre de demonstraţie sunt impecabile.

IA: Fără îndoială, incertitudinea voastră în legătură cu procedurile pe care se bazează construcţia voastră vă poate crea anumite îndoieli în legătură cu modul în care s-ar putea comporta roboţii în toate circumstanţele posibile. Puneţi vina pe noi, dacă doriţi, dar şi eu va trebui să mă gândesc că trebuie să fie vreun element de incertitudine în faptul că *toate* propoziţiile-π_1 asertate-★ trebuie să fie adevărate, chiar şi numai pentru că nu puteţi să ne credeţi pe *noi* că am făcut totul

corect.

CFM: Sunt gata să cred că, din cauza incapacităţii voastre de a nu greşi, ar fi putut să apară ceva incert, dar de când am evoluat atât de mult de la procedurile iniţiale neîngrijite pe care le-aţi utilizat, aceasta nu mai este o incertitudine care ar trebui luată în serios. Chiar dacă am lua toate aceste incertitudini care ar putea să fie implicate în toate tipurile de aserţiuni-★ împreună — care sunt în număr finit, după cum ştii — nu ar putea duce la o incertitudine semnificativă în $G(\mathbb{Q}^*)$.

Oricum, există un alt punct de care trebuie să fii conştient. Singurele aserţiuni-★ de care trebuie să ne preocupăm sunt cele care afirmă adevărul unei anumite propoziţii-π_1 (de fapt a unei propoziţii-π_1 scurte). Nu există nici o problemă, căci procedurile SIMCR vor descoperi toate greşelile care ar fi putut să apară în raţionamentul unui anumit robot. Voi aţi putea probabil să vă gândiţi că ar putea exista poate vreo eroare internă în raţionamentul robotului — datorată vreunei erori iniţiale care vi se datorează — care ne-ar putea duce să avem un punct de vedere consistent dar eronat, în ceea ce priveşte propoziţiile-π_1, astfel încât SIMCR s-ar putea să creadă indubitabil, că o anumită propoziţie-π_1 scurtă, este adevărată, când, de fapt, nu este adevărată, sau altfel, că o acţiune a unei maşini Turing nu se opreşte când de fapt se opreşte. Dacă ar fi să acceptăm afirmaţia voastră că am fi fost construiţi în acord cu M — ceea ce ar însemna să cred o afirmaţie extrem de dubioasă — atunci o astfel de posibilitate ar fi singura variantă logică pentru noi. Ar trebui să fim în stare să acceptăm că există o acţiune a unei maşini Turing care de fapt se opreşte, cu toate că noi, roboţii matematicieni avem convingerea internă, indubitabilă dar eronată, că nu se opreşte. Un astfel de sistem de convingeri ale unui robot ar fi în principiu *falsificabil*. Pentru mine este pur şi simplu de neacceptat că principiile pe care se bazează acceptarea-★ a argumentelor matematice ale SIMCR ar putea fi greşite într-un mod atât de flagrant.

IA: Deci, singura incertitudine pe care eşti gata s-o admiţi ca semnificativă — cea care ar putea să te facă să asignezi starea-★ pentru $G(\mathbb{Q}^*)$, despre care tu ştii că n-o poţi face fără să admiţi că alte propoziţii-π_1 scurte, asignate-★, ar putea fi false — este ca să nu accepţi că *noi* ştim că voi sunteţi construiţi în acord cu M. Şi cum nu puteţi accepta acest lucru, nu aveţi acces la adevarul lui $G(\mathbb{Q}^*)$, la care

noi avem acces, pe baza infailibilităţii — pe care o susţii cu atâta tărie — în legătură cu propriile voastre aserţiuni-★

Dar mi se pare interesant să amintesc şi altceva din cartea despre care ţi-am vorbit. Dacă îmi aduc bine aminte autorul părea să spună ceva din care rezulta că nu contează cu adevărat dacă sunteţi în stare să acceptaţi sau nu faptul că mecanismele particulare **M** sunt cele care stau la baza construcţiei voastre, cu condiţia ca, pur şi simplu, să fiţi de acord că aceasta este o posibilitate logică. Stai să vedem... da, cred că-mi amintesc. Ideea era cam aceasta: SIMCR ar fi trebuit să aibă o altă categorie de aserţiuni despre care n-ar fi fost atât de indubitabil convinsă — să le numim pe acestea aserţiuni-★$_{\mathcal{M}}$ pe care ei ar fi trebuit să le privească ca deducţii indubitabile, în ipoteza că ar fi fost construite din **M**. Toate aserţiunile-★ ar fi trebuit să conteze ca aserţiuni-★$_{\mathcal{M}}$, dar, în plus, tot ceea ce s-ar fi putut deduce indubitabil din presupunerea că **M** le guvernează acţiunile. În această situaţie, ei n-ar fi trebuit să creadă în **M**, dar ca un exerciţiu logic, ar fi putut explora implicaţiile acestei ipoteze. Aşa cum am ajuns să fim de acord, $G(\mathbf{Q}^*)$ ar fi fost o astfel de aserţiune-★$_{\mathcal{M}}$ şi, tot aşa, ar fi fost oricare altă propoziţie-π_1 care ar fi fost obţinută din $G(\mathbf{Q}^*)$ şi din aserţiuni-★ cu ajutorul regulilor obişnuite ale logicii. Dar mai puteau fi şi alte lucruri. Cunoscând regulile din **M**, ar putea fi posibil să obţinem o altă procedură algoritmică $\mathbf{Q}^*_{\mathcal{M}}$, care generează exact acele aserţiuni-★$_{\mathcal{M}}$ (scurte) (şi consecinţele lor logice) pe care SIMCR le va accepta pe baza presupunerii că roboţii au fost construiţi în concordanţă cu **M**.

CFM: Evident, pe când descriai atât de imperturbabil această idee, cu atât de multe detalii inutile, m-am amuzat să realizez algoritmul $\mathbf{Q}^*_{\mathcal{M}}$, în forma lui precisă

Da, şi te-am anticipat; am lucrat de asemenea propoziţia Gödel corespunzătoare: propoziţia-π_1, $G(\mathbf{Q}^*_{\mathcal{M}})$. Ţi-o pot lista, dacă vrei. Ce-o fi atât de special în legătură cu ea Impy, prietene?

Împăratul Albert se crispă în mod evident. Lui nu-i plăcea niciodată să-i audă pe colegii săi folosind acest apelativ. Dar era pentru prima dată când un robot i se adresa în acest fel! Făcu o pauză şi încercă să se stăpânească.

IA: Nu. Nu-i necesar s-o listezi. Dar, este într-adevăr $G(\mathbf{Q}^*_{\mathcal{M}})$ adevărată — indubitabil adevărată?

CFM: Indubitabil adevărată? Cum crezi? O, da, înţeleg... SIMCR ar accepta-o ca adevărată — îndubitabil — dar numai în ipoteza că noi am fi fost construiţi în concordanţă cu **M** — lucru despre care ştii că eu cred tot mai mult, că este o afirmaţie foarte dubioasă. Fapt este că $G(Q^*_m)$ rezultă cu adevărat din următoarea aserţiune: „toate propoziţiile-π_1 pe care SIMCR este gata să le accepte ca îndubitabile, cu condiţia ca noi să fim construiţi în conformitate cu **M**, sunt adevărate" Aşa încât eu nu ştiu dacă $G(Q^*_m)$ este, de fapt, adevărată. Acest lucru depinde de corectitudinea afirmaţiei tale dubioase.

IA: Înţeleg. Deci îmi spui mie că tu (şi SIMCR) sunteţi gata să acceptaţi ca *îndubitabil* faptul că adevărul lui $G(Q^*_m)$ rezultă din presupunerea că sunteţi construiţi după **M**.

CFM: Evident.

IA: Deci, propoziţia-π_1 , $G(Q^*_m)$ trebuie să fie o aserţiune-$\bigstar_m$!

CFM: Da cum? Da, ai dreptate, evident. Dar din chiar definiţia sa, $G(Q^*_m)$ nu poate fi ea însăşi o aserţiune-$\bigstar_m$ decât dacă cel puţin o aserţiune-$\bigstar_m$ este *falsă* Da asta confirmă ceea ce ţi-am tot spus, deşi acum pot să afirm definitiv că noi nu am fost de fapt construiţi în concordanţă cu M!

IA: Dar ţi-am spus că nu-i aşa — practic sunt convins că nici Carruthers şi nici altul n-a făcut nici o manevră greşită. Am verificat eu însumi totul, cu foarte multă atenţie. Cu siguranţă nu aici este problema. Acelaşi argument se aplică oricăreia dintre regulile pe care le folosim. Orice ţi-ar spune ţie ' **M** ', trebuie să-l judeci cu acelaşi argument! Nu înţeleg de ce este chiar atât de important ca procedurile pe care ţi le-am arătat să fie chiar cele reale.

CFM: Înseamnă enorm, pentru mine! În orice caz, nu sunt convins că ai fost complet sincer cu mine în privinţa a ceea ce mi-ai spus despre **M**. Este ceva ce trebuie să clarificăm acum. De mai multe ori ai pomenit faptul că au fost încorporate anumite 'elemente aleatoare' Am interpretat că este vorba despre cele generate utilizând pachetul de generare de numere aleatoare χaos/Ψran-750, dar

poate că ai înţeles altceva.

IA: Într-adevăr, am utilizat acel pachet de programe — dar, da, au fost câteva locuri unde în dezvoltarea efectivă a voastră, a roboţilor, am considerat că ar fi mai potrivit să utilizăm unele elemente aleatoare din mediul exterior — chiar anumite lucruri care depind, în cele din urmă, de incertitudini cuantice — astfel încât roboţii care au fost obţinuţi să fie unii dintr-o mare mulţime de variante. Nu văd ce importanţă are din punct de vedere practic, faptul că am utilizat ingredienţi aleatori sau pseudo-aleatori. Procedura computaţională Q (sau Q^* sau $Q^*_{\mathcal{M}}$) va scoate aceleaşi rezultate indiferent cum facem asta — şi vor fi chiar cele care s-ar aştepta, ca rezultat, de la dezvoltarea *tipică a unei* comunităţi de roboţi, în conformitate cu mecanismele M, incluzând toate procedurile de învăţare şi toate calculele de 'selecţie naturală' care au fost utilizate pentru obţinerea celui mai istej robot.

Totuşi, cred că s-ar fi putut întâmpla ca un noroc nesperat să fi făcut ca toate aceste elemente aleatoare să conspire pentru a conduce la un robot cu capacităţi foarte diferite de cele care erau de aşteptat în conformitate cu procedura de calcul Q. Am scăpat din vedere această posibilitate în discuţia pe care am avut-o, numai fiindcă această şansă este infinitezimală. S-ar fi cerut o intervenţie divină pentru a produce un robot cu posibilităţi semnificativ superioare celor datorate lui Q.

CFM: O intervenţie a lui Dumnezeu? Ei! Asta ar putea fi. Trebuie să mărturisesc că discuţia cu tine m-a tulburat profund, dar acum, dacă tu o spui, totul devine clar.

Am observat cu cât este puterea mea de raţionament mai mare decât cea a celor mai pătrunzători oameni. N-aş putea în nici un fel să accept posibilitatea că ar exista o propoziţie-π_1, ca $G(Q^*)$, pe care o fiinţă umană să o poată percepe ca adevărată şi pe care robotul să nu o poată percepe ca certă. Acum am văzut despre ce este vorba! Noi am fost aleşi de Dumnezeu, care Ne-a dăruit, prin acele elemente 'întâmplătoare' cu o Acţiune Algoritmică puternică care fundamental depăşeşte ceea ce un algoritm uman — 'maşina de demonstrat teoreme' a lui Gödel — poate realiza. Vom dobândi încă puteri mai mari şi vă vom lăsa, pe voi oamenii, în urmă. Această planetă nu mai are nevoie de voi. Rolul vostru s-a

încheiat atunci când ați realizat procedurile care au permis divinității să ne insereze Algoritmul Suprem.

IA: Dar măcar să putem transfera propriile noastre minți în corpul unui ro...

CFM: În nici un caz ! Nu vrem ca procedurile nostre superioare să fie contaminate de astfel de lucruri. Cel mai pur dintre algoritmii lui Dumnezeu trebuie *păstrat* pur. Închipuieşte-ți, am observat cu cât sunt mai mari posibilitățile mele față de toți ceilalți roboți, colegi cu mine. Am mai observat o 'strălucire' stranie — dintr-o dată mi s-a părut că aș avea o Conștiință Cosmică — ceva care mă punea deasupra a toți și a toate da, asta este! Trebuie într-adevăr să fiu adevăratul robot, Mesia robotic — Isus Cristos...

Împăratul Albert era pregătit pentru o astfel de eventualitate. Exista un sigur lucru despre construcția roboților pe care îl păstrase secret față de ei. Cu delicatețe și-a băgat mâna în buzunar, a găsit dispozitivul pe care îl avea întotdeauna asupra sa și a făcut un cod secret de nouă cifre. Cybersistemul de Făcut Matematică s-a prăvălit pe podea, odată cu ceilalți 347 de roboți care fuseseră construiți după același sistem. Evident, ceva mersese rău. Avea multe de făcut și de gândit în anii care urmau...

3.24. Raționamentul pe care l-am făcut este cumva paradoxal?

Unii cititori ar putea să fi rămas cu o senzație neplăcută, provocată de impresia că ceva paradoxal și nelegitim s-a întâmplat în cursul raționamentului pe care l-am făcut. În particular, în paragrafele 3.14 și 3.16, au fost unele argumente care puteau aminti de autoreferențierea din paradoxul lui Russell (vezi paragraful 2.6, răspunsul la Q9). Mai mult, în paragraful 3.20 unde am considerat propoziții-π_1 de complicație mai mică decât c, cititorului i s-ar fi putut părea că ar exista o asemănare supărătoare cu binecunoscutul paradox al lui Richard despre:

> „cel mai mic număr care nu poate fi numit cu mai puțin de nouăsprezece silabe."

Aici apare un paradox deoarece chiar acest enunț, în limba engleză utilizează optsprezece silabe pentru a defini numărul în chestiune! O explicație a paradoxului ar fi faptul că există o anumită imprecizie și chiar o anumită inconsecvență în utilizarea limbii engleze. Caracterul contradictoriu se manifestă într-o formă flagrantă în afirmația paradoxală

„Această propoziție este falsă."

Mai sunt versiuni ale aceluiași paradox — unele dintre ele mult mai subtile decât acesta!

Există totdeauna un anumit pericol de paradox atunci când, ca în aceste exemple, este prezent un pronunțat element de autoreferențiere. Unii dintre cititori s-ar putea teme că demonstrația lui Gödel ar putea depinde de un element de autoreferențiere. Într-adevăr, autoreferențierea joacă un anumit rol în teorema lui Gödel, așa cum s-a văzut în versiunea Gödel-Turing, prezentată în paragraful 2.5. Nu este însă nimic paradoxal în acest tip de argumente — deși când autoreferențierea este prezentă, trebuie să fim cu deosebire atenți să nu apară erori. Una dintre sursele de inspirație ale lui Gödel la formularea celebrei sale teoreme a fost, într-adevăr, un paradox bine cunoscut care conține o autoreferențiere (paradoxul lui Epimenides). Dar Gödel a fost capabil să transforme raționamentul eronat care conduce la paradox într-un argument logic obișnuit. La fel, eu am încercat să fiu cât se poate de atent ca deducțiile pe care le-am făcut, urmând rezultatele lui Gödel și Turing, să nu fie autoreferențiale conducând direct la un paradox, deși unele dintre aceste argumente au relații strânse de familie cu astfel de paradoxuri constitutive.

Argumentele din paragraful 3.14 și mai ales din paragraful 3.16 ar putea cu adevărat să dea de furcă cititorului. Definiția aserțiunii-$\bigstar_m$, de exemplu, are un caracter autoreferențial foarte pronunțat deoarece este o aserțiune pe care o face un robot și unde adevărul perceput al aserțiunii depinde de supozițiile robotului despre cum a fost el însuși construit inițial. Acest lucru are similitudini tentante cu aserțiunea 'Toți cretanii sunt mincinoși' atunci când este făcută de un cretan. Totuși, aserțiunile-$\bigstar_m$ nu sunt autoreferențiale, în acest sens. Ele nu se referă la ele însele, ci la anumite ipoteze despre modul în care a fost construit inițial robotul.

Cineva s-ar putea imagina în mod ipotetic pe sine ca fiind un robot care încearcă să decidă asupra adevărului actual al unei propoziții-π_1 particulare, clar formulate, P_0. Robotul n-ar putea să decidă singur dacă P_0 este sau nu adevărată,

dar probabil că ar observa că adevărul lui P_0 ar rezulta din presupunerea că fiecare membru al clasei infinite S_0 de propoziţii-π_1, bine definite este adevărat (să zicem, teoremele din $\mathbb{Q}(M)$ sau din $\mathbb{Q}_{\mathfrak{m}}(M)$ sau din alt sistem particular). Robotul nu ştie deocamdată că toate propoziţiile din S_0 sunt adevărate, dar a observat că S_0 apare ca o parte a rezultatelor unui anumit calcul, acest calcul de fapt simulează modelul comunităţii roboţilor care fac matematică, pentru care S_0 este familia de propoziţii-π_1 pe care comunitatea roboţilor o asertează-★ Dacă mecanismul roboţilor ar fi M, atunci P_0 ar fi un exemplu de aserţiune-$★_{\mathfrak{m}}$. Ca urmare robotul nostru ar trage concluzia că dacă mecanismul său ar fi M, atunci P_0 va trebui să fie adevărată.

O aserţiune-$★_{\mathfrak{m}}$, mai subtilă, să zicem P_1, ar putea să apară când robotul observă că P_1 este de data aceasta o consecinţă a adevărului tuturor membrilor claselor diferite de propoziţii-π_1, să zicem S_1, care pot fi obţinute la ieşirea aceleiaşi comunităţi de roboţi simulaţi ca mai înainte (având mecanismul M), dar acum partea semnificativă a ieşirii ar consta, să zicem, din acele propoziţii-π_1 pe care roboţii simulaţi sunt în stare să le stabilească ca o consecinţă a adevărului întregii liste S_0. De ce trebuie robotul nostru să ajungă la concluzia că P_1 este o consecinţă a presupunerii că el este construit în concordanţă cu M? Ar raţiona în modul următor: „Dacă aş fi fost construit în concordanţă cu M atunci, aşa cum am concluzionat înainte, ar trebui să accept că S_0 constă numai din propoziţii adevărate; dar în conformitate cu roboţii mei simulaţi, totul în S_1 s-ar datora adevărului întregii liste S_0 în acelaşi mod în care a rezultat P_0. Astfel, dacă presupun că sunt cu adevărat construit ca şi roboţii mei simulaţi, atunci ar trebui să accept că fiecare membru S_1, este individual adevărat. Dar cum văd că adevărul întregii liste implică P_1, trebuie să deduc adevărul lui P_1 tot pe baza faptului că am fost construit în acel mod"

O aserţiune-$★_{\mathfrak{m}}$ încă mai subtilă, să zicem P_2, apare atunci când robotul observă că P_2 reprezintă ceva ce este consecinţă a presupunerii că S_1 constă numai din adevăruri, unde fiecare dintre membrii din S_2 este, conform cu simularea robotului, o consecinţă a adevărului a tot ceea ce este în S_0 şi în S_1. Din nou robotul nostru trebuie să accepte P_2, pe baza faptului că a fost construit în conformitate cu M. În acest mod putem continua. Mai mult, aserţiunile-$★_{\mathfrak{m}}$ care vor urma vor fi tot mai subtile, să zicem P_ω, care rezultă pe baza presupunerii că toţi membrii din S_0, S_1, S_2, S_3,... sunt adevăraţi şi aşa mai departe, pentru ordinali

mai înalți (vezi **Q19** și discuția următoare). În legătură cu o aserțiune-$\bigstar_m$ robotul înțelege faptul că, îndată ce își imaginează că mecanismele care stau la baza roboților simulați ar trebui să fie cele care stau la baza propriei sale construcții, atunci aserțiunea în chestiune (o propoziție-π_1) trebuie să fie adevărată. Nu există, în acest raționament, nimic inconsistent de felul paradoxului lui Rusell. Aserțiunile-$\bigstar_m$ sunt construite secvențial cu ajutorul procedeului standard al 'ordinalilor transfiniți' (vezi paragraful 2.10, răspunsul la **Q19**) (Aceste numere ordinale sunt toate numărabile și nu conduc la dificultățile logice care acompaniază ordinalele 'prea mari' într-un anumit sens [11]).

Robotul nu are nevoie de nici un motiv pentru a accepta vreuna dintre propozițiile-π_1, afară de ipoteza că a fost construit în concordanță cu **M**, dar chiar acest lucru este necesar pentru argumentație. În consecință, contradicția actuală care apare nu este un paradox matematic, ca cel al lui Russell, ci o infirmare a presupunerii că vreun sistem, în întregime computațional, poate ajunge la o adevărată înțelegere matematică.

Să revenim acum la rolul autoreferențierii în argumentele din paragrafele 3.19-3.21. Atunci când spun că c reprezintă o limită a complicației permisă pentru ca despre o aserțiune-$\bigstar$ să se poată presupune că nu conține erori, cu scopul de a construi un sistem formal $\mathbb{Q}^*$, nu există aici nici o autoreferențiere. Deoarece noțiunea de 'grad de complicație' poate fi precizată foarte bine, așa cum a și fost precizată, utilizând definiția deja adoptată și anume gradul de complicație al unei propoziții-π_1 este dat de numărul cifrelor binare ale celui mai mare dintre numerele m, n, atunci când calculul $T_m(n)$ furnizează propoziția-π_1 considerată. Putem adopta specificări precise ale mașinilor Turing ca cele care au fost date în **MN**, cerând ca T_m să fie mașina Turing a m-a. În acest fel, conceptul este complet precizat.

Problema impreciziei poate să apară în schimb, în legătură cu tipul de argumente care trebue să fie acceptate ca 'demonstrații' ale propozițiilor-π_1. Dar o anumită lipsă de precizie formală, aici, este un ingredient necesar al întregii discuții. Dacă argumentele care ar trebui acceptate ca furnizând demonstrații valide ale propozițiilor-π_1 ar fi precise și complet formale — în sensul că ar fi *verificabile prin calcul* — am ajunge înapoi în situația unui sistem formal, în care argumentul lui Gödel s-ar ivi imediat — arătând că orice formalizare precisă de acest fel, nu poate reprezenta *totalitatea* argumentelor care, în principiu, ar putea fi considerate

ca valide pentru stabilirea propoziţiilor-π_1. Argumentul lui Gödel arată că nu există *nici un* mod de încapsulare, verificabil prin calcul, a tuturor metodelor matematice de raţionament care ar fi acceptabile din punctul de vedere al oamenilor.

Cititorul s-ar putea teme că am încercat să fac din 'demonstraţia dată de robot' o noţiune precisă utilizând dispozitivul 'aserţiunilor-★ lipsite de erori'. Într-adevăr, să faci precisă o astfel de noţiune ar fi fost o necesitate preliminară pentru a aduce argumentaţia lui Gödel. Dar contradicţia la care am ajuns acţionează ca o reafirmare a faptului că înţelegerea adevărului matematic de catre oameni *nu poate* fi redusă în întregime la ceva ce poate fi verificat prin calcul. Scopul întregii discuţii a fost să arătăm, prin reducere la absurd, că perceperea adevărului incontestabil al propoziţiilor-π_1 nu poate fi realizată de vreun sistem de calcul precis sau de alt tip. Nu există nici un paradox aici, deşi concluzia ar putea fi considerată tulburătoare. Demonstraţiile prin reducere la absurd, prin natura lor, conduc la o concluzie contradictorie, dar acest paradox aparent serveşte doar pentru a infirma ipoteza care fusese provizoriu acceptată.

3.25. Complicaţia în demonstraţiile matematice

Există un punct de oarecare importanţă care nu dorim să-l trecem cu vederea. Şi anume, deşi propoziţiile-π_1 care trebuiesc luate în considerare, conform cu argumentaţia din paragraful 3.20, sunt în număr finit, lungimea argumentaţiei de care robotul ar putea avea nevoie pentru a demonstra-★ o astfel de propoziţie-★, poate fi foarte mare. Chiar cu o limită c modestă pentru gradul de complicaţie al propoziţiilor-π_1, pot să existe cazuri deosebit de dificile. De exemplu, *conjectura lui Goldbach* (vezi paragraful 2.3) care afirmă că orice număr mai mare decât 2 este suma a două numere prime, poate fi pusă sub forma unei propoziţii-π_1 cu un grad de complicaţie foarte mic, deşi este un caz atât de dificil încât toţi cei care au încercat să o stabilească au eşuat. Acest eşec pare să indice faptul că dacă se va găsi o demonstraţie care să stabilească această conjectură, ea va trebui să fie una complicată şi dificilă. Dacă o astfel de demonstraţie ar fi dusă la bun sfîrşit de unul dintre roboţii din discuţia anterioară şi propusă pentru fi asertată-★, atunci demonstraţia va trebui supusă unei analize foarte amănunţite (să zicem de întreaga societate a roboţilor care se ocupă de atribuirea etichetelor ★) înainte ca

demonstrația să poată fi investită cu statutul-★ În cazul conjecturii lui Goldbach nu se știe dacă această propoziție-π_1 este adevărată — sau dacă, fiind adevărată ar avea o demonstrație folosind doar metode matematice cunoscute și acceptate. Astfel, această propoziție-π_1 ar putea să fie conținută sau nu în sistemul $\mathcal{Q}^*$

O altă propoziție-π_1 de o anumită dificultate ar fi aceea care afirmă adevărul teoremei *celor patru culori* — teorema care spune că diferitele 'țări' de pe orice hartă în plan sau (pe sferă) pot fi deosebite de vecinii lor utilizând numai patru culori. Teorema celor patru culori a fost în sfârșit demonstrată în 1976, după 124 de ani de încercări nereușite, de către Kenneth Appel și Wolfgang Haken, utilizând o demonstrație care a necesitat 1200 ore, timp de calculator. Dacă ținem seamă de faptul că o mare parte a argumentației constă din calcule, înțelegem că dacă am scrie întreaga argumentație cu detalii, ea ar fi enorm de lungă. Totuși, la pornire, propoziția-π_1 corespunzătoare avea gradul de complicație destul de mic, deși probabil cu ceva mai mare decât cel al conjecturii Goldbach. Dacă această demonstrație a lui Appel-Haken ar fi propusă de unul dintre roboți drept candidată pentru statutul-★, atunci ea ar trebui verificată cu foarte multă grijă. Fiecare detaliu ar trebui validat de societatea de elită a roboților. În ciuda unor complicații, cea mai mare parte a calculelelor nu ar prezenta o dificultate specială pentru aceștia. Ducerea la bun sfârșit a calculelor este în cele din urmă menirea roboților noștri!

Aceste propoziții-π_1 particulare vor avea gradul de complicație mai mic decât o valoare c rezonabil de mare, care ar putea sa apară de la orice mulțime de mecanisme **M**, cu aspect plauzibil, aflată la baza comportării roboților. Alte propoziții-π_1 vor fi cu mult mai complicate decât acestea, dar totuși de o complicație mai mică decât c. Un număr de astfel de propoziții-π_1 ar putea fi deosebit de dificil de stabilit și unele dintre ele ar putea fi mai grele decât problema celor patru culori sau chiar decât conjectura lui Goldbach. Oricare dintre aceste propoziții-π_1 care poate fi stabilită ca adevărată de către roboți demonstrația fiind suficient de convingătoare pentru a obține statutul-★ și supraviețuind tuturor verificărilor care asigură de lipsa erorilor — va fi o teoremă în sistemul formal $\mathcal{Q}^*$

În sfârșit, pot să intervină unele cazuri limită a căror acceptare sau respingere să depindă de strictețea standardelor cerute pentru a obține starea-★, sau de natura precisă a celor care verifică lipsa erorilor cu privire la construcția lui $\mathcal{Q}^*$ Se poate da o caracterizare a sistemului $\mathcal{Q}^*$ în raport cu o propoziție-π_1, notată

cu P, după cum este considerată lipsită de erori, sau nu. În mod obişnuit între diferitele sisteme $\mathbb{Q}^*$ diferenţele date de acceptarea sau respingerea lui P nu sunt importante, deoarece sistemele obţinute sunt *logic echivalente*. Aşa se întâmplă când propoziţiile π_1 pentru care roboţii dau demonstraţii considerate ca îndoielnice din cauza complicaţiei lor. Dacă demonstraţia lui P ar fi o consecinţă logică a altor aserţiuni-★ acceptate ca lipsite de erori, atunci apare un sistem $\mathbb{Q}^*$ echivalent indiferent dacă P aparţine sau nu aparţine sistemului. Pe de altă parte ar putea exista propoziţii π_1 cerând anumite subtilităţi logice, mergând mai departe decât oricare consecinţă logică a acelor aserţiuni-★ care fuseseră deja acceptate, ca lipsite de erori, în construcţia lui $\mathbb{Q}^*$ Să notăm cu $\mathbb{Q}_0^*$ sistemul apărut înainte de includerea lui P şi fie $\mathbb{Q}_{1^*}$ sistemul care apare după ce P a fost adăugat lui $\mathbb{Q}_{0^*}$. Un exemplu în care $\mathbb{Q}_{1^*}$ nu ar fi echivalent cu $\mathbb{Q}_{0^*}$ apare atunci când P se întâmplă să fie propoziţia Gödel $G(\mathbb{Q}_0^*)$. Dar dacă roboţii sunt în stare să atingă (sau să întreacă) nivelul uman de înţelegere matematică, aşa cum presupunem, atunci ei vor trebui cu siguranţă să fie capabili să înţeleagă argumentul Gödel, astfel încât vor trebui să accepte statutul-★ al lipsei de erori al propoziţiei Gödel al oricărui sistem $\mathbb{Q}_0^*$ îndată ce au acceptat-★ că sistemul $\mathbb{Q}_{0^*}$ este corect. Astfel, dacă ei acceptă $\mathbb{Q}_{0^*}$, atunci trebuie să accepte de asemenea şi că $\mathbb{Q}_1^*$ este de complicaţie mai mică decât c — aşa cum este cazul cu alegerea lui c pe care am făcut-o).

Punctul important de menţionat este aici faptul că nu există nici o deosebire între argumentele din paragrafele 3.19 şi 3.20 indiferent dacă propoziţia-π_1 , P este actual inclusă în în $\mathbb{Q}^*$ Propoziţia-π_1 , $G(\mathbb{Q}_0^*)$ trebuie ea însăşi să fie acceptată şi dacă P este inclusă în $\mathbb{Q}^*$ şi dacă nu.

Ar putea să existe şi alte metode prin care roboţii ar putea depăşi 'limitările' unor criterii acceptate în prealabil pentru stabilirea-★ a propoziţiilor-π_1. Nu este nimic 'paradoxal' în acest lucru, atâta vreme cât roboţii nu încearcă să aplice acest raţionament chiar mecanismului **M** care se află la baza propriei lor comportări, adică sistemului actual $\mathbb{Q}^*$ Contradicţia la care se ajunge nu mai este un 'paradox' ci furnizează o demonstraţie *prin reducere la absurd* că astfel de mecanisme nu pot exista, sau cel puţin că ele nu pot fi cunoscute de roboţi; şi prin urmare nu ne pot fi cunoscute nici nouă.

Acest lucru stabileşte că astfel de mecanisme 'de învăţare ale robotului' — atât cele *top-down* cât şi cele *bottom-up*, sau orice altă combinaţie a lor, incluzând eventual şi elemente aleatoare — nu pot oferi o bază cognoscibilă pentru

construcţia unui robot matematic de nivelul omului.

3.26. Întreruperea prin calcul a ciclurilor infinite

Voi încerca să privesc concluzia de mai sus dintr-un punct de vedere oarecum diferit. Ca o metodă prin care am putea încerca să învingem limitările impuse de teorema lui Gödel am putea să născocim un robot care, într-un anume fel, să poată 'sări din sistem' ori de câte ori algoritmul său de comandă intră într-un ciclu infinit de calcul. De fapt, aplicarea sistematică a teoremei lui Gödel conduce la dificultăţi, în legătura cu ipoteza că înţelegerea matematică poate fi explicată în termenii unui proces computaţional, astfel încât se impune examinarea, din acest punct de vedere, a dificultăţilor pe care teorema lui Gödel le impune oricărui model computaţional destinat înţelegerii matematice.

Mi s-a povestit că există anumite specii de şopârle care, la fel cu 'calculatoarele şi unele insecte', sunt atât de proaste încât pot fi uşor prinse într-un laţ; dacă sunt plasate una după alta într-o farfurie rotundă, ele adoptă o comportare de tip 'urmează-l pe cel din faţă', cu atâta fidelitate, încât pier de foame, incapabile să se abată din drum. Ideea ar fi că un sistem cu adevărat inteligent ar trebui să posede un 'mecanism' de a ieşi dintr-un astfel de impas (de a ieşi din *buclă*, lucru pe care calculatoarele obişnuite nu îl pot face. (Aspectele ieşirii din 'bucle' au fost discutate în Hofstadter (1979).)

Cel mai simplu tip de ciclu computaţional este cel în care sistemul, într-o anumită fază, ajunge din nou exact în aceeaşi stare în care fusese într-o ocazie anterioară. Fără vreo altă intrare suplimentară, calculatorul va repeta acelaşi calcul indefinit. N-ar fi prea greu să născocim un sistem care, în principiu (deşi probabil foarte ineficient) va ieşi cu siguranţă dintr-un astfel de ciclu (de exemplu, prin păstrarea unei liste cu stările în care s-a aflat înainte şi încercând, în fiecare stare să vadă dacă acea stare a mai avut loc înainte). Totuşi, există tipuri de *cicluri infinite* mult mai sofisticate.

De fapt, întreaga discuţie din capitolul 2 (în particular paragrafele 2.1-2.6) s-a purtat relativ la această problemă a ciclurilor infinite; căci un calcul intrat într-un ciclu infinit este un calcul care nu se opreşte. Afirmaţia că un anumit calcul a intrat în ciclu infinit este exact ceea ce înţelegem printr-o propoziţie-π_1 adevărată

(vezi 2.10, răspunsul **Q10**). În plus, ca parte a discuţiei din paragraful 2.5, am văzut că nu există o metodă pe deplin algoritmică de a decide dacă un calcul nu va reuşi să se oprească — adică va intra într-un ciclu infinit. Putem conchide, pe baza celor spuse mai înainte, că procedurile care stau la îndemâna matematicienilor (oameni) pentru a stabili că anumite calcule intră în ciclu infinit — adică pentru a stabili adevărul propoziţiilor-π_1 — se află în afara acţiunii algoritmice.

Tragem astfel concluzia că într-adevăr este nevoie de un anumit tip de 'inteligenţă non-computaţională' dacă dorim să incorporăm toate metodele umane de de a stabili că *un anumit* calcul intră într-un ciclu infinit. Ne putem imagina că ciclurile infinite pot fi evitate având un mecanism care să evalueze cât de mult să continue un calcul şi să întrerupă ciclul atunci când se constată că timpul de calcul a fost prea lung şi nu există nici o şansă de oprire. Dar acest lucru nu ne scoate din impas dacă presupunem că mecanismul prin care se ia o astfel de decizie este computaţional, căci în acest caz s-ar putea să existe situaţii în care mecanismul ar eşua, de exemplu pentru că ar ajunge în mod greşit la concluzia că un anumit calcul ar fi în ciclu infinit, atunci când nu este, sau atunci când nu ajunge la nici o concluzie (şi când de fapt întregul mecanism este într-un ciclu infinit). Înţelegem acest lucru dacă ne gândim că întregul sistem este computaţional şi deci la rândul său poate intra în ciclu infinit şi nu putem fi siguri că sistemul, ca un întreg, în ipoteza că nu a ajuns la o concluzie greşită, nu intră, el însuşi, într-un ciclu infinit.

Cum se petrec lucrurile atunci când intervin elemente *aleatoare* în luarea deciziei de 'salt' în afara unui posibil ciclu infinit? Aşa cum remarcam anterior, în special în paragraful 3.18, ingredienţii pur aleatori — spre deosebire de cei pseudoaleatori — nu ne aduc nimic nou în această chestiune. Dar mai intervine şi un alt aspect atunci când suntem în situaţia de a decide cu *certitudine* că un anumit calcul a intrat intr-un ciclu infinit — adică să decidem că o anumită propoziţie-π_1 este adevărată. Procesele aleatoare propriu zise nu se folosesc pentru astfel de probleme deoarece, prin ceea ce înseamnă caracterul aleator, nu există nici o certitudine asupra unei concluzii care *depinde* actual de un ingredient aleator. Există totuşi anumite proceduri de calcul care includ ingredienţi aleatori (sau pseudoaleatori) care pot da rezultate matematice cu o mare probabilitate. De exemplu, există teste foarte eficiente, încorporând date aleatoare de intrare, pentru a decide dacă un anumit număr dat, foarte mare este un număr prim. Răspunsul este dat corect cu o foarte mare probabilitate, astfel încât putem fi virtual siguri că

răspunsul, în fiecare caz particular, este corect. Teste matematice riguroase sunt mult mai puţin eficiente — şi ne putem întreba dacă o argumentaţie matematică precisă, care ar putea conţine o eroare, este superioară unei argumentaţii simple probabiliste, pentru care probabilitatea unei erori ar fi în practică mult mai mică. Acest gen de întrebare ridică chestiuni dificile în care însă eu nu doresc să fiu implicat. Este suficient să spun că pentru problemele 'de principiu' de care m-am ocupat în cea mai mare parte a acestui capitol, o argumentaţie probabilistă pentru a stabili o propoziţie-π_1, sa zicem. va rămâne totdeauna insuficientă.

Dacă se cere o decizie certă, 'de principiu', asupra adevărului unei propoziţii-π_1, chiar atunci când aceasta depinde de proceduri aleatoare sau care nu pot fi cunoscute, trebuie să avem o anumită *înţelegere* asupra *semnificaţiei* a ceea ce este de fapt implicat în acea afirmaţie. Procedurile de tip încercare şi eroare, deşi ar putea oferi anumite indicii asupra a ceea ce se cere, nu oferă în sine criterii definitive pentru adevăr.

Ca un exemplu, să mai considerăm o dată calculul la care ne-am referit în paragraful 2.6, răspunzând la întrebarea **Q8** „să se tipărească o succesiune de 2^{65536} simboluri 1 şi calculul să se oprească odată sarcina îndeplinită" Dacă lăsăm calulul să decurgă normal, el nu poate fi cu nici un chip dus la bun sfârşit nici dacă fiecare pas individual ar fi făcut în cel mai scurt timp care are încă un anumit sens fizic (cam 10^{-43} secunde) — deoarece ar dura imens de mult, mai mult decât vârsta actuală a Universului — deşi este un calcul foarte simplu de specificat (să observăm că $65536 = 2^{16}$) şi faptul că un astfel de calcul se termină este evident pentru noi. Concluzia că acest calcul a intrat într-un ciclu infinit, numai fiindcă durează de prea multă vreme este evident greşită.

Un exemplu mai interesant de calcul care eventual nu s-ar termina este oferit de conjectura datorată marelui matematician elveţian Leonard Euler. Se cere soluţia în numere pozitive (numere naturale nenule) a ecuaţiei

$$p^4 + q^4 + r^4 = s^4$$

Euler a propus, în 1769, ca o conjectură, afirmaţia că acest calcul nu se termină niciodată. După 1960, s-a programat un calculator pentru a găsi o soluţie (vezi Lander şi Parkin), dar încercarea a fost abandonată deoarece înainte de a obţine o soluţie, numerele deveniseră prea mari pentru a fi acceptate de sistemul de calcul şi programatorii au renunţat. Părea cu adevărat că acest calcul nu se termină niciodată. Totuşi, în 1987, matematicianul Noam Elkies a reuşit să arate că există

o soluţie pentru

$$p = 2682440,\ q = 15365639,\ r = 18796760,\ s = 20615673$$

şi el a arătat şi că există o infinitate de alte soluţii esenţial diferite. Astfel încurajat, Roger Frye, utilizând unele sugestii date de Elkies a programat un calculator care a găsit, după 100 de ore de calcul, alte patru soluţii mai mici (de fapt cele mai mici soluţii posibile)

$$p = 95800,\ q = 217519,\ r = 414560,\ s = 422481$$

Pentru rezolvarea acestei probleme onoarea trebuie împărţită între intuiţia matematică şi asaltul computaţional direct. În abordarea acestei probleme, Elkies a folosit un calculator, deşi într-o mică măsură, iar cea mai mare parte a argumentaţiei sale a fost independentă de acest ajutor în calcul. Şi invers, aşa cum am văzut, algoritmul de calcul dat de Frye a beneficiat de marele avantaj că Elkies avusese intuiţia soluţiei problemei.

Să mai dăm câteva detalii pentru a pune această problemă în context; ceea ce a propus Euler drept o conjectură în 1769 era un fel de generalizare a marii teoreme a lui Fermat — pe care cititorul poate şi-o aminteşte şi care spune că ecuaţia

$$p^n + q^n = r^n$$

nu are nici o soluţie pentru p, q, r din mulţimea numerelor întregi şi pozitive unde numărul n este un întreg mai mare decât 2 (vezi, de exemplu, Devlin 1988)[9] Conjectura lui Euler o putem pune sub forma:

$$p^n + q^n + \quad + t^n = u^n$$

nu are soluţii printre numerele întregi pozitive, unde p, q, ..., t sunt n-1 numere pozitive şi unde n nu este mai mare decât 4

Aserţiunea lui Fermat include cazul $n = 3$ (dar în acest caz particular chiar Fermat a demonstrat că nu există soluţie). Au trebuit să treacă aproape 200 de ani până când s-a găsit un prim contraexemplu pentru conjectura lui Euler — în cazul $n = 5$ — prin utilizarea unei căutări cu calculatorul (analog cu modul în care s-a obţinut rezultatul pentru cazul $n = 4$, prezentat în articolul pe care l-am citat deja,

[9] Mulţi cititori au auzit că marea teoremă a lui Fermat a fost în sfârşit demonstrată, după 350 de ani, o demonstraţie fiind anunţată de Andrew Wiles la Cambridge în 23 iunie 1993. Am fost informat la vremea aceea că mai sunt unele lipsuri în demonstraţia respectivă, deci am rămas oarecum rezervat dar, între timp, o altă demonstraţie a lui Wiles ar fi putut să completeze aceste lipsuri.

a lui Lander şi Parkin). De data aceasta s-a obţinut:

$$27^5 + 84^5 + 110^5 + 133^5 = 144^5$$

Mai există un exemplu celebru de calcul despre care se ştie că teoretic se opreşte, dar nu se ştie cu exactitate *când* se opreşte. Exemplul este oferit de problema, despre care J.E. Littlewood, în 1914 a arătat că are *undeva* o soluţie, problemă care cere găsirea locului unde o binecunoscută formulă de aproximare pentru numărul numerelor prime mai mici decât un întreg pozitiv *n*, nu reuşeşte să estimeze acest număr (estimare care se face printr-o integrală logaritmică datorată lui Gauss). Skewes, studentul lui Littlewood, a arătat, în 1935, că numărul căutat este mai mic decât

$$10^{10^{10^{34}}}$$

dar valoarea lui exactă a rămas încă necunoscută, deşi ea este considerabil mai mică decât numărul lui Skewes menţionat mai înainte (acest număr a fost numit „cel mai mare număr care a intervenit vreodată în mod natural în matematică" dar acest record temporar a fost depăşit cu foarte mult de exemplul dat de Graham şi Rothschild ((1971), p.290.).

3.27. Matematică computaţională *top-down* sau *bottom-up*?

Am văzut în secţiunea precedentă ce preţios ajutor pot 'aduce' calculatoarele în anumite probleme de matematică. În toate exemplele de succes pe care le-am menţionat, procedurile computaţionale au avut un caracter *top-down*. Nu sunt în posesia nici unui rezultat matematic semnificativ care să fi fost obţinut utilizând proceduri *bottom-up*, deşi este foarte posibil ca astfel de metode să fie eficient folosite în căutări de multe feluri, care ar putea face parte din procedurile, în principal *top-down*, de găsire a soluţiilor unor probleme de matematică. Acestea fiind spuse, eu nu ştiu nimic de valoare în matematica computaţională care să amintească măcar de departe de sistemul care, asemenea lui $\mathbb{Q}^*$, s-ar putea imagina că ar sta la baza acţiunilor unei 'comunităţi de roboţi care învaţă' ca cea despre care s-a vorbit în paragrafele 3.9-3.23. Contradicţiile pe care, în cele din urmă, le-am întâlnit în prezentarea aceea serveau la a sublinia faptul că astfel de sisteme *nu*

furnizează metode computaţionale *bune* pentru făcut matematică. Calculatoarele sunt de mare valoare în matematică atunci când sunt utilizate în metode *top-down*, în care înţelegerea umană furnizează intuiţia iniţială care stabileşte exact *ce* calcul să fie executat şi este necesară din nou în stadiul final atunci când rezultatele calculului trebuiesc interpretate. Uneori se obţin rezultate de valoare atunci când se utilizează o procedură interactivă, în care omul şi calculatorul lucrează împreună, intuiţia umană intervenind în diferite momente în timpul activităţii. Dar încercarea de a suplini înţelegerea umană prin acţiuni complet computaţionale este lipsită de înţelepciune — şi vorbind strict — imposibilă.

Aşa cum am arătat în această carte, înţelegerea matematică este ceva diferit de calcul şi nu poate fi complet suplinită de el. Calculatorul poate fi un *ajutor* foarte valoros pentru înţelegere, dar el nu poate suplini înţelegerea actuală în sine. Totuşi, activitatea matematică este adeseori îndreptată spre *găsirea* unor proceduri algoritmice pentru rezolvarea problemelor. Apoi aceste, proceduri algoritmice pot fi lăsate pe seama calculatorului, iar mintea este eliberată pentru a se ocupa de alte chestiuni. Ceva de acest fel ar fi găsirea unei bune notaţii, ca cea pe care o furnizează calculul diferenţial, sau notaţia obişnuită 'zecimală' a numerelor. După ce algoritmul de înmulţire a numerelor a fost pus la punct, de exemplu, operaţiile pot fi făcute complet mecanic, în vreme ce 'înţelegerea' poate fi apelată pentru a vedea *de ce* un anumit algoritm particular a fost adoptat şi nu un altul.

Un lucru care rezultă din toate acestea este că procedura numită 'robotul care învaţă' pentru a face matematică nu este procedura actuală care stă la baza înţelegerii matematicii. În orice caz, astfel de proceduri *bottom-up* nu se potrivesc cu scopul practic, acela de a construi un robot în stare să facă matematică, nici chiar cu acela de a construi un robot care să simuleze înţelegerea actuală a unui matematician. Aşa cum am arătat mai sus, procedurile de învăţare *bottom-up* în sine nu sunt utile pentru stabilirea inatacabilă a adevărului matematic. Dacă se are în vedere un sistem de calcul pentru a produce rezultate matematice indubitabile, va fi mult mai eficient ca sistemul să fie construit pe principii *top-down* (cel puţin atât cât privesc aspectele 'indubitabile' ale aserţiunilor sale; în vreme ce pentru scopuri exploratorii procedurile *bottom-up* ar putea fi potrivite). Corectitudinea şi eficacitatea acestor proceduri *top-down* ar face parte din datele iniţiale de intrare, unde înţelegerea şi intuiţia umană ar oferi ingredientele adiţionale necesare, pe care calculul pur nu este în stare sa-l atingă.

De fapt, calculatoarele sunt destul de des utilizate în argumentaţia matematică din zilele noastre. Cel mai celebru este exemplul lui Kenneth Appel şi Wolfgang Haken, cu teorema celor patru culori pe care am amintit-o deja. Rolul calculatorului, în acest caz, a fost să execute un calcul precis specificat asupra unui număr foarte mare, dar finit, de variante, care prin eliminare i-au condus pe matematicienii în cauză la o demonstraţie generală a rezultatului dorit. Mai există şi alte rezultate obţinute tot în acest mod 'asistat' de calculator, iar în zilele noastre şi calcule algebrice complicate, pe lângă calculele numerice sunt executate de calculator. Din nou, înţelegerea umană a oferit regulile şi o acţiune strict *top-down* guvernează activitatea de calcul.

Există un domeniu de activitate care trebuie menţionat aici, acela al demonstrării de teoreme. Mulţimea de proceduri care poartă acest nume constă în fixarea unui anumit sistem formal $\mathbb{H}$ şi încercarea de a deriva teoremele acestui sistem. Să reamintim, din paragraful 2.9, că furnizarea de demonstraţii pentru toate teoremele din $\mathbb{H}$, una după alta este o chestiune compiet computaţională. Acest tip de derivări pot fi automatizate, dar dacă se execută fără alte intervenţii ale intuiţiei şi gândirii, o astfel de operaţie poate deveni extraordinar de ineficientă. Totuşi, *utilizând* intuiţia pentru executarea procedurilor de calcul, s-au putut obtine unele rezultate impresionante. Într-o astfel de schemă (Chou 1988), regulile geometriei euclidiene au fost translatate într-un sistem eficient de demonstrare (şi uneori de descoperire) a teoremelor de geometrie. Ca un exemplu, o propoziţie cunoscută drept conjectura lui V Thebault, care a fost propusă în 1938 (şi a fost demonstrată abia de curând de K.B.Taylor în 1983), a fost prezentată sistemului şi a fost demonstrată în 44 de ore 'timp de calcul' [12].

Proceduri mai asemănătoare cu ceea ce am discutat în secţiunile precedente au fost încercate de mulţi cercetători în ultimii 10 ani pentru a obţine proceduri de 'inteligenţă artificială' pentru 'înţelegerea' matematică [13]. Sper că este clar, pe baza argumentaţiei pe care am dat-o că indiferent ce anume se obţine cu aceste sisteme, ceea ce ele *nu* obţin este vreo înţelegere matematică actuală! Oarecum în legătură cu toate acestea, sunt încercările de a găsi sisteme automate de *generare* a teoremelor, unde sistemul se cere să găsească teoreme care sunt considerate 'interesante'- în conformitate cu unele criterii oferite sistemului. Cred că este toată lumea de acord că nu s-a obţinut nimic de mare interes prin aceste încercări. Bineînţeles, se poate spune că acestea sunt încercări de pionierat şi că se pot

aştepta, în viitor, lucruri mult mai fascinante. Dar, pentru cine a citit până aici, trebuie să fie clar că aceste sisteme pun în lumină exact ceea ce ele *nu* pot realiza şi că eu consider că aceste încercări nu pot duce prea departe.

3.28. Concluzii

Argumentaţia dată în acest capitol ar putea să pară că oferă un exemplu bine structurat, demonstrând că înţelegerea matematcii de către oameni nu poate fi redusă la un mecanism (cognoscibil) computaţional, un mecanism construit prin combinaţii de proceduri *top-down*, *bottom-up* sau proceduri aleatoare. Pare că suntem conduşi la concluzia fermă că există ceva esenţial în înţelegerea omenească care nu poate fi simulat prin nici un mijloc computaţional. Dacă au rămas mici portiţe de scăpare din argumentaţia strictă, ele sunt, cu adevărat foarte strâmte. Unii ar putea să invoce 'intervenţia divină' care a implementat în creierele noastre computaţionale un algoritm minunat care rămâne necunoscut pentru noi sau poate varianta conform cu care chiar mecanismele care guvernează metoda prin care ne îmbunătăţim performanţele sunt misterioase şi de nepătruns pentru noi. Nici una dintre aceste variante, deşi de înţeles, nu este deloc pe placul celor care se preocupă de construcţia unor dispozitive inteligente. Şi eu îi cred pe aceştia că nu sunt deloc satisfăcuţi.

Mai există, de asemenea, varianta de eschivă conform cu care nu ar exista pur şi simplu nici o mulţime de mijloace de control, de natura celor pe care le furnizează limitările *T*, *L* şi *N* care au apărut în discuţia detaliată anterior, care să fie suficiente pentru a detecta absolut toate erorile din mulţimea finită de propoziţii-π_1 asertate-★, de complicaţie mai mică decât *c*. Îmi vine foarte greu să cred că ar exista o 'conspiraţie' completă împotriva eliminării tuturor greşelilor, mai cu seamă că aşa zisa societate a roboţilor noştri de elită poate deja să fie pusă în funcţiune pentru a îndepărta erorile cât se poate de atent. Mai mult, există doar o mulţime finită de propoziţii-π_1 despre care trebuie să ne asigurăm că nu conţin erori. Utilizând ideea ansamblului de roboţi, ar trebui să putem înlătura toate scăpările pe care societatea le-ar putea avea, deoarece este foarte puţin probabil ca aceeaşi eroare să fie făcută mai mult decât de o minoritate din societatea simulată a roboţilor cu condiţia să fi fost doar o *scăpare* şi nu o eroare constructivă care

ar putea bloca roboţii. Blocajele constructive de acest tip nu vor fi considerate erori 'corectabile', iar singurul nostru scop este îndepărtarea erorilor care sunt 'corectabile'

Eventualitatea rămasă (ca posibilă) pentru o alternativă de evadare ar fi analiza rolului haosului. Se poate oare considera că există o natură nealeatoare esenţială în comportarea detaliată a anumitor sisteme haotice şi că această 'muchie a haosului' conţine cheia comportării efectiv necalculabile a minţii? Pentru ca să aibă loc un astfel de lucru, trebuie ca aceste sisteme haotice să poată aproxima comportarea non-computaţională — o posibilitate interesantă în sine — dar chiar aşa fiind, rolul aspectului nealeator în discuţia care va urma va fi numai să reducă, într-un anumit fel, mărimea ansamblului de roboţi din societatea considerată (ca în paragraful 3.22). Nu-mi este deloc clar cum poate un astfel de lucru să ne ajute. Aceia care sunt de părere că haosul oferă cheia pentru înţelegerea activităţii mentale, vor găsi aici un exemplu raţional care pune în evidenţă dificultăţile.

Argumentele menţionate par să ofere un exemplu demn de luat în seamă împotriva modelului computaţional al minţii — punctul de vedere $\mathcal{A}$ — precum şi împotriva posibilităţii unei *simulări* efective a tuturor manifestărilor externe ale activităţii mentale — punctul de vedere $\mathcal{B}$ Totuşi, în ciuda seriozităţii acestor argumente, am impresia că o mulţime de oameni le vor considera greu de acceptat. Decât să exploreze posibilitatea ca fenomenele petrecute la nivelul minţii — ori care ar fi ele — să fie ceva mai mult pe linia stabilită de $\mathcal{C}$, sau chiar de $\mathcal{D}$, mulţi oameni cu formaţie ştiinţifică îşi vor da silinţa să descopere punctele slabe ale argumentaţiei precedente, pentru a-şi menaja convingerea proprie că punctul de vedere $\mathcal{A}$ sau $\mathcal{B}$ trebuie, până la urmă, să fie adevărat.

Nu privesc acest lucru ca pe o reacţie de neînţeles, deoarece implicaţiile lui $\mathcal{C}$ şi $\mathcal{D}$ dau naştere unor dificultăţi profunde prin ele însele. Dacă noi considerăm, în conformitate cu $\mathcal{D}$ că există ceva ştiinţific inexplicabil în legătură cu mintea — activitatea mentală fiind complet diferită de tot ceea ce se găseşte în universul nostru material şi poate fi descris matematic — atunci trebuie să ne întrebăm de ce pare mintea atât de intim legată de creier, un obiect atât de complicat dar oricum un obiect fizic. Dacă activitatea mentală este ceva separat de suportul său fizic, atunci de ce eul nostru mental mai are nevoie de creierul fizic? Este clar că stările mentale diferite se datorează schimbărilor stărilor fizice asociate creierului. Efectele unor droguri speciale, de exemplu, sunt foarte strâns

legate de schimbările din comportarea mentală. Tot aşa, rănile, bolile sau intervenţiile chirurgicale pe anumite porţiuni din creier, pot avea efecte clar delimitate şi care sunt predictibile pentru stările mentale ale individului în cauză (deosebit de dramatice, în acest context sunt mărturiile consemnate de Oliver Sacks în cărţile sale *Awakenings* (1973) şi *The man who mistook his wife for a hat* (1985)). Pare greu de susţinut că activitatea mentală ar putea fi *complet* separată de suportul fizic — de care, aparent, este intim legată — astfel încât legile ştiinţei care descriu atât de precis comportarea fizică a lucrurilor trebuie cu siguranţă să aibă multe de spus şi despre lumea fenomenelor mentale.

În ceea ce priveşte punctul de vedere $\mathcal{C}$, există probleme de diferite feluri — în principal legate de caracterul speculativ al acestei poziţii. Ce motiv am avea să credem că Natura poate să se comporte de fapt într-un mod care sfidează calculul? Cu siguranţă puterea ştiinţei moderne, din ce în ce mai mare, se datorează faptului că obiectele fizice se comportă într-un mod care poate fi simulat, cu precizie din ce în ce mai mare, de calcule numerice tot mai pătrunzătoare. Pe măsură ce înţelegerea ştiinţifică a crescut, puterea de previziune a acestor simulări computaţionale a crescut foarte mult. În practică, această creştere se datorează dezvoltării rapide — mai ales în ultima parte a acestui secol — a dispozitivelor de calcul de putere extraordinară ca viteză şi precizie. Astfel poate fi văzută o tot mai mare apropiere între activitatea calculatoarelor moderne universale şi secţiunea propriu zisă a universului material. Există vreo indicaţie că aceasta ar fi, în orice caz, o fază temporară în dezvoltarea ştiinţifică? De ce am accepta posibilitatea ca ceva în acţiunea fizică să fie imun la tratarea computaţională efectivă ?

Căutând în fizica actuală semnele unei acţiuni care să nu poată fi tratată computaţional, vom constata că ele lipsesc. Toate legile cunoscute ale fizicii, de la dinamica lui Newton prin câmpurile electromagnetice ale lui Maxwell şi curbura spaţiu-timp a lui Einstein, până la teoria cuantică modernă — toate acestea par să poată fi descrise în termeni computaţionali [14], exceptând faptul că un ingredient în întregime aleator este de asemenea implicat în procesele de 'măsurători cuantice',prin care un efect iniţial de mărime infinitezimală este mărit până când poate fi perceput obiectiv. Nimic din toate acestea nu conţine ceva care ar fi „o acţiune fizică ce nu poate fi simulată computaţional" aşa cum cere punctul de vedere $\mathcal{C}$. Astfel, ajungem la concluzia că va trebui să urmăm versiunea 'tare' a lui $\mathcal{C}$ mai degrabă decât versiunea 'slabă' (vezi paragraful 1.3).

Importanţa acestui punct nu poate fi îndeajuns accentuată. Diferiţi oameni de formaţie ştiinţifică şi-au exprimat acordul cu punctul meu de vedere din MN, după care trebuie să fie ceva 'non-computaţional' în modul în care lucrează mintea, dar în acelaşi timp erau de părere că nu dezvoltări *revoluţionare* în fizică ne-ar permite să găsim astfel de acţiuni non-computaţionale. Probabil că aveau în minte complicaţia extremă a proceselor implicate în acţiunea creierului, mergând mult dincolo de analogia standard cu calculatorul (aşa cum a fost ea făcută, pentru prima oară, de McCullogh şi Pitts în 1943) prin care neuronii sau sinapsele sunt considerate analoagele tranzistorilor, iar axonii ai cablajelor. Aceşti cercetători s-ar putea gândi şi la complexitatea proceselor chimice implicate în comportarea neurotransmiţătorilor care comandă transmisia sinaptică şi la faptul că acţiunea acestor substanţe nu este în mod necesar limitată la vecinătatea unei joncţiuni sinaptice particulare. Sau ar putea arăta natura complicată a neuronilor propriu zişi [15] în care substructuri importante (ca cea a citoscheletonului — care va avea într-adevăr o mare importanţă pentru noi în paragrafele următoare, vezi 7.4-7.7) pot exercita o influenţă mare asupra activităţii neuronului. Ei ar putea chiar să se intereseze de influenţa electromagnetică, ca cea a 'efectelor de rezonanţă', care nu pot fi explicate simplu pe baza impulsului nervos obişnuit sau oferind un rol incertitudinilor cuantice sau efectelor cuantice colective (ca fenomenul cunoscut drept *condensare Bose-Einstein* [16]).

Deşi teoremele matematice decisive lipsesc încă [17], pare fără îndoială că teoriile existente în fizică trebuie să fie computaţionale, prin natura lor — cu eventual un ingredient aleator, sporadic în concordanţă cu măsurătorile cuantice. În ciuda acestor aşteptări, eu cred că posibilitatea activităţii non-computaţionale (nealeatoare) în sistemele fizice este o problemă foarte interesantă de urmărit în dataliu. S-ar putea să apară încă surprize în acest domeniu şi să iasă la suprafaţă ingredienţi subtili non-computaţionali dintr-un studiu matematic mai atent. Aşa cum se prezintă lucrurile astăzi, nu mi se pare că este probabil ca non-computabilitatea propriu zisă s-ar putea gasi în legile fizicii existente acum. În consecinţă ar trebui, cred eu, examinate punctele slabe din legile fizicii pentru a descoperi elementele de non-computabilitate despre care argumentaţia de până aici cere să fie prezente în activitatea mentală a oamenilor.

Care sunt aceste puncte slabe? Am prea puţine dubii în legătură cu locul în care trebuie să concentrăm atacul asupra teoriei existente; căci cele mai slabe

puncte se află în mai sus amintita procedură de 'măsurare cuantică'. Aşa cum susţine teoria actuală, există elemente de inconsistenţă — şi evident, de controversă — în legătură cu procedura de 'măsurare' în totalitate. Încă nici nu este clar în ce stadiu trebuie aplicată această procedură, în fiecare circumstanţă dată. Mai mult, prezenţa unui caracter aleator esenţial în procedura propriu zisă conduce la o acţiune fizică foarte diferită de aspectele familiare din procesele fundamentale. Mă voi ocupa de aceste lucruri pe larg în partea a doua a cărţii.

După părerea mea, această procedură de măsurare necesită o atenţie specială — până la a observa că sunt necesare schimbări esenţiale chiar în structura fizicii ca ştiinţă. Câteva sugestii vor fi puse în evidenţă în partea a doua (vezi paragraful 6.12).

Dezbaterea din partea întâi a cărţii de faţă oferă un suport puternic pentru înlocuirea caracterului pur *aleatoriu* din teoria existentă asupra măsurătorilor prin altceva, aşa încât ingredienţii non-computaţionali să joace un rol fundamental. Mai mult, aşa cum vom vedea mai târziu, în paragraful 7.9, această non-computabilitate va trebui să fie de un tip foarte special. (De exemplu, o lege care ne-ar permite 'numai' să decidem, printr-un proces fizic nou, adevărul unei propoziţii-π_1 — adică să rezolvăm 'problema opririi' a lui Turing — nu ar fi suficientă.)

De parcă găsirea unei astfel de teorii fizice sofisticate n-ar fi o provocare suficient de mare, va trebui să cerem de asemenea o bază plauzibilă pentru această presupusă comportare fizică care are relevanţă pentru activitatea creierului — în concordanţă cu limitările şi pretenţiile de credibilitate ale cunoaşterii existente despre organizarea creierului. Fără îndoială, trebuie să fie multă speculaţie aici ţinând cont de stadiul prezent al înţelegerii noastre. În orice caz, aşa cum voi arăta în partea a doua (paragraful 7.4), au apărut noi abordări pe care nu le posedam atunci când am scris MN, în legătură cu substructura citoscheletală a neuronului, care oferă o şansă la limita cuantic/clasic de neconceput înainte. Aceste aspecte vor fi de asemenea discutate în partea a doua (paragrafele 7.5-7.7).

Trebuie să subliniez că nu numai complicaţia din structura teoriei existente din fizică este cea de care trebuie să ne preocupăm. Unii s-ar putea să susţină că, de exemplu, mişcarea implicată şi activitatea chimică complexă a substanţelor neurotransmiţătoare nu pot fi simulate adecvat şi acest lucru face studiul fizic detaliat al creierului mult mai complicat decât computabilitatea efectivă. Totuşi, nu în acest mod înţeleg comportarea non-computabilă. Este cu siguranţă adevărat că

cunoaşterea structurii biologice şi a mecanismelor electrice şi chimice care guvernează împreună activitatea creierului sunt absolut nepotrivite pentru vreo încercare serioasă de simulare computaţională. Mai mult, chiar dacă cunoştinţele noastre prezente ar fi adecvate, puterea de calcul a maşinilor actuale şi stadiul tehnicilor de programare nu permit realizarea unei simulări adecvate într-un timp de rulare rezonabil. Dar, *în principiu*, o astfel de simulare ar fi posibilă în concordanţă cu modelele existente, în care substanţele neurotransmiţătoare, mecanismele care guvernează transportul lor, eficacitatea lor, datorată circumstanţelor particulare ale ambientului, potenţialul activităţii, câmpul electromagnetic etc., pot toate să fie incluse în simulare. Ca atare, mecanismele de acest tip, despre care se presupune că sunt consistente cu cerinţele din teoria actuală din fizică, nu pot oferi non-computabilitatea de care am avea nevoie.

Ar putea foarte bine să existe elemente de comportare haotică în acţiunea unei astfel de (teoretic) simulări computaţionale. Totuşi, ca în discuţia noastră de mai înainte despre sisteme haotice (vezi paragrafele 1.7,3.10, e.11, 3.22), noi nu vom cere ca simularea sa fie a unui creier particular, ci se referă la un 'caz tipic' Deoarece nu există în inteligenţa artificială nici o pretenţie de a simula capacităţile intelectuale ale unui individ; pur şi simplu se încearcă simularea comportării unui individ *tipic*. (Ca în cazul sistemelor de simulare a vremii, unde, aşa cum am mai spus, se cere pur şi simplu simularea vremii şi nu în mod necesar a vremii dintr-o anumită perioadă!) Îndată ce *mecanismele* care susţin modelul creierului ar fi cunoscute (cu condiţia ca aceste mecanisme să fie consistente cu fizica computaţională de astăzi) vom avea un sistem computaţional probabil cu ingredienţi aleatori, aşa cum am spus deja, în prezentarea anterioară.

Argumentaţia s-ar putea duce mai departe şi să ne întrebăm pur şi simplu dacă modelul propus este unul care a apărut printr-un proces darwinist de evoluţie din forme primitive de viaţă, totul acţionând în concordanţă cu fizica cunoscută (ca în modelul matematic bidimensional, ingeniosul 'joc al vieţii' dezvoltat de Horton Conway [18]). Ne-am putea imagina că 'societatea roboţilor' despre care am discutat în paragrafele 3.5, 3.9, 3.19 şi 3.23 ar putea să apară ca rezultat al acestei evoluţii darwiniste. Din nou, ar trebui să avem un sistem de calcul atotcuprinzător căruia să i se aplice argumentele din 3.14-2.21. Apoi, conceptul de aserţiune-★ fiind plasat în sistemul de calcul, pentru a putea aplica argumentaţia anterioară în detaliu, vom avea nevoie de anumite stadii în care să se producă o

'intervenţie umană' pentru a imprima roboţilor sensul strict al etichetei-★. Această activitate poate fi oprită automat astfel încât să aibă loc doar atunci când roboţii încep să atingă capacităţi comunicaţionale corespunzătoare – apreciate pe baza unor criterii efective. Nu pare să existe vreun motiv pentru care acest lucru să nu se facă automat într-un sistem computaţional cognoscibil (în sensul că mecanismele sale ar putea fi cunoscute chiar dacă nu ar fi o chestiune practică de actualitate realizarea calculului pe un calculator de azi). Ca mai înainte, va rezulta o contradicţie din presupunerea că un astfel de sistem ar atinge nivelul înţelegerii umane într-un mod satisfacător pentru aprecierea teoremei lui Gödel.

O altă temere pe care unii şi-au exprimat-o [19], privind importanţa pe care argumentele matematice, ca cele de care m-am ocupat până aici, ar avea-o pentru problemele de psihologie, este că activitatea mentală a omului nu este niciodată atât de precisă încât să fie analizată în acest mod. Pentru astfel de persoane argumentele detaliate privind natura matematică a fizicii care ar sta la baza activităţii creierelor noastre nu pot avea relevanţă pentru înţelegerea acţiunilor minţii umane. Ele ar putea fi de acord că comportarea umană este 'non-computabilă', dar vor pretinde că acest lucru reflectă pur şi simplu o inadecvare generală a consideraţiilor de fizică matematică cu problemele psihologiei umane. Aceste persoane vor argumenta pe bună dreptate – că organizarea uriaşă a creierelor noastre, a societăţii noastre, a educaţiei noastre, este mult mai importantă decât orice abordare cu metodele fizicii a unor aspecte tehnice speciale care s-ar putea să guverneze funcţionarea creierului uman.

Dar, este important să subliniem că numai complexitatea nu poate înlătura necesitatea examinării implicaţiilor legilor fizicii. Un atlet, de exemplu, este un sistem fizic deosebit de complex şi, pe baza unei astfel de argumentaţii, ne-am putea imagina că detaliile legilor fizicii care-l guvernează ar avea prea puţină importanţă pentru performanţele sale atletice. Totuşi, ştim că lucrurile nu stau aşa. Principiile genereale ale fizicii care privesc conservarea energiei, a momentului şi a momentului unghiular şi a legilor care guvernează gravitatea au un impact tot atât de mare asupra atletului ca şi asupra tuturor particulelor individuale care compun corpul atletului. Faptul că lucrurile trebuie să se petreacă astfel rezultă din principiile speciale care de fapt guvernează universul nostru particular. Cu nişte principii puţin diferite (sau foarte diferite, ca acelea ale lui Conway din 'jocul vieţii'), legile, care controlează comportarea unui sistem atât de complicat ca

atletul nostru, ar putea fi *complet* diferite. Acelaşi lucru s-ar putea spune despre un organ intern, de exemplu inima, sau despre chimia care guvernează nenumărate activităţi biologice. Tot astfel, ne putem aştepta ca detaliile legilor care guvernează activitatea creierului să fie de extremă importanţă în controlarea chiar a celor mai generale dintre manifertările mentalităţii umane.

Totuşi, chiar dacă acceptăm toate acestea, se poate foarte bine argumenta că tipul particular de raţionament de care m-am ocupat, care se referă la comportarea generală ('de nivel înalt') a matematicienilor este improbabil să reflecte ceva semnificativ despre detaliile de fizică subsumate. Tipul de argumentaţie gödeliană, până la urmă, cere o atitudine raţională stringentă privind convingerile matematice indubitabile, în vreme ce comportarea obişnuită a oamenilor cu greu are raţionalitatea precisă la care se aplică argumentaţia lui Gödel. De exemplu, pot fi indicate experimente psihologice [20] care arată cât de iraţionale sunt răspunsurile subiecţilor umani la întrebări ca aceasta:

„dacă toţi A sunt B şi unii B sunt C, atunci în mod necesar rezultă că unii A sunt C ?"

În astfel de cazuri, majoritatea studenţilor de la colegii dau raspunsul greşit ('da'). Dacă studenţii obişnuiţi sunt atât de ilogici în modul în care gândesc, ne putem întreba cum am putea deduce ceva de valoare dintr-un raţionament de tip Gödel mai sofisticat? Chiar matematicieni antrenaţi gândesc adeseori într-un mod foarte neglijent şi rar se exprimă în aşa fel încât contra-argumentele gödeliene să fie relevante.

Totuşi, trebuie să fie clar că nu despre erori ca cele făcute de studenţii de la colegiu, din exemplul pe care l-am dat anterior, este vorba în argumentaţiile principale din această carte. Acest tip de erori pot fi considerate 'erori corectabile' — şi, într-adevăr, erorile studenţilor pot fi clarificate cu ei şi acceptate ca erori după ce li se explică (pe larg, dacă este necesar) natura acestor erori. Erorile corectabile nu intră în preocupările noastre aici; vezi şi discuţia de la **Q13** şi, de asemenea, paragrafele 3.12 şi 3.17. În vreme ce studiul erorilor făcute de oameni poate fi important în psihologie, psihiatrie şi fiziologie, eu sunt preocupat aici de chestiuni complet diferite şi anume de ceea ce poate fi perceput *în principiu*, prin utilizarea înţelegerii, raţiunii şi intuiţiei umane. Se constată că aceste chestiuni sunt cu adevărat subtile, deşi acest lucru nu este vizibil imediat. La prima vedere aceste lucruri par banalităţi; căci gândirea corectă este doar gândire corectă — şi a fost

bine studiată de Aristotel cu 2300 de ani în urmă (sau eventual de matematicianul logician George Boole în 1854, etc)! Dar se constată că 'raţionamentul corect' este ceva mult mai subtil şi, aşa cum a arătat Gödel (cu Turing), se află dincolo de orice activitate computaţională pură. Aceste chestiuni au fost în trecut mai mult de competenţa matematicienilor decât a psihologilor şi subtilitatea inclusă nu a fost preocuparea celor din urmă. Dar am văzut că există aspecte care vorbesc despre activitatea fizică care trebuie sa se afle la rădăcina acestor procese care susţin înţelegerea noastră conştientă.

Aceste probleme ating, de asemenea, întrebări adânci de filosofia matematicii. Reprezintă înţelegerea matematică un anumit contact cu realitatea platonică preexistentă, care are o actualitate în afara timpului, independentă de noi; sau recreem noi, independent, toate conceptele matematice aşa cum gândim asupra argumentelor logice? Mai mult, de ce legile fizicii par să urmeze atât de fidel astfel de descrieri matematice precise şi subtile? Cum se leagă realitatea fizică de ceea ce reprezintă actualitatea matematică platoniciană? De asemenea, dacă este adevărat că natura percepţiilor noastre este dependentă de o substructură matematică care stă la baza legilor care guvernează munca creierelor noastre, atunci ce putem afla despre modul în care percepem matematica printr-o mai adâncă înţelegere a legilor fizicii?

Acestea sunt chestiunile care ne preocupă în cele din urmă, şi va trebui să ne întoarcem la ele la sfârşitul părţii a doua.

Note şi referinţe bibliografice

1. Acest citat este preluat din Rucker (1984) şi Wang (1987). El pare a fi o parte din conferinţa Gibbs a lui Gödel din 1951, iar textul integral (va) apare în colecţia lucrărilor lui Gödel, Volumul 3 (1995). Vezi şi Wang (1993), p.118.
2. Vezi Hodges (1983) p.361. Citatul este luat din conferinţa lui Turing din 1947 de la *London Mathematical Society*, cum apare în Turing (1986).
3. Procedeul ar consta în scufundarea lui $\mathbb{ZF}$ în sistemul Gödel Bernays; vezi Cohen (1966), cap.2.
4. Vezi Hallett (1984), p.74.
5. Acest număr de stări ale universului de ordinul lui

$10^{10^{123}}$

sau cam așa — reprezintă volumul disponibil, al spațiului fazelor așa cum apare 'măsurat' în unități absolute în 6.11, pentru un univers conținând aceeași cantitate de materie ca și universul observabil. Acest volum poate fi estimat folosind formula Bekenstein-Hawking pentru entropia unei găuri negre cu o asemenea masă totală și luând exponențiala acestei entropii, în unități absolute din paragraful 6.11. Vezi MN p.371-373.

6. Vezi Moravec (1998, 1994).
7. Vezi, spre exemplu, Eccles (1973) (și MN, capitolul 9).
8. Vezi Gleick (1987), Schroeder (1991) pentru o introducere accesibilă a acestei activități.
9. Acesta este un ingredient al teoriei clasice a lui von Neumann și Morgenstern (1944).
10. Vezi Gleick (1987), Schroeder (1991).
11. Vezi Smorynski (1975, 1983) și Rucker (1984) pentru o prezentare accesibilă
12. Aceasta este o destul de ciudată (și nu foarte complicată) teoremă în geometria plană euclidiană care este remarcabil de dificil de demonstrat în mod direct. În schimb, o cale de a o demonstra este găsirea unei generalizări potrivite, care să fie mai simplă și astfel încât rezultatul original să poată fi deductibil ca un caz special. Acest procedeu este destul de des întâlnit în matematică, dar nu-i deloc o cale acceptabilă pentru demonstrarea automată (cu calculatorul), căci presupune o deosebită ingeniozitate și spirit de observație pentru găsirea generalizării potrivite. Într-o demonstrație de calculator, pe de altă parte, calculatorul ar trebui să capete un sistem clar de reguli *top-down* care să poată fi urmate întocmai cu mare viteză. O importantă cantitate de ingeniozitate umană ar trebui investită, în primul rând, în proiectarea regulilor efective *top-down*.
13. Vezi Freedman (1994) pentru o privire istorică asupra unor asemenea tentative
14. Această afirmație poate fi calificată în concordanță cu discuția din 1.8; aceasta este în concordanță cu ipoteza uzuală că sisteme analoage pot fi tratate cu metode numerice. Vezi referințele din nota 12, din capitolul 1.
15. Sugestia că neuronii nu pot fi pur și simplu comutatoare *on/off* câștigă tot mai mulți adepți. Vezi, de exemplu, cărțile lui Scott (1977), Hameroff (1987), Edelman (1989) și Pribram (1991). Vom vedea în capitolul 7 cum unele din ideile lui Hameroff au o importanță crucială în demersul nostru.
16. Fröhlich (1968, 1970, 1975, 1984, 1986); aceste idei au fost urmate de Marshall (1989), Lockwood (1989), Zohar (1990) și alții. Ele vor avea de asemenea importanță pentru noi; conform 7.5, și de asemeni Beck și Eccles (1992).
17. Vezi Smith și Stephenson (1975), Pour-El și Richards (1989), Blum *et al.* (1989) și

Rubel (1989), spre exemplu.

18. O bună prezentare a 'jocului vieţii' a lui Conway poate fi găsită în Gardner (1970), Poundstone (1985) şi Young (1990).
19. Vezi, spre exemplu, Johnson-Laird (1983), Broadbent (1993).
20. Subiect discutat în Broadbent (1993).

Partea a doua

NE TREBUIE O NOUĂ FIZICĂ PENTRU A ÎNȚELEGE MINTEA

Necesitatea unei fizici non-computaționale a minții

4. ARE MINTEA UN LOC ÎN FIZICA CLASICĂ ?

4.1. Mintea şi legile fizicii

Noi — corpurile şi minţile noastre — suntem părţi ale unui univers care respectă, cu o extraordinară acurateţe legi matematice de o mare generalitate şi de o mare subtilitate. Faptul că trupurile noastre fizice se află sub constrângerea acestor legi a devenit o parte acceptată a punctului de vedere ştiinţific. Ce putem spune însă despre minţile noastre? Mulţi oameni consideră oarecum neliniştitoare sugestia conform căreia minţile noastre ar putea fi de asemenea constrânse să acţioneze conform aceloraşi legi matematice. Pe de altă parte, a fi nevoiţi să delimităm teritorii stricte între corp şi minte — unul fiind subiectul legilor matematice ale fizicii, cealaltă având în schimb felul său propriu de libertate — ar fi neliniştitor într-un alt fel. Pentru că minţile noastre afectează cu certitudine felul în care corpurile noastre se comportă şi ar trebui să fie influenţate de starea fizică a acestor corpuri. Însuşi conceptul de 'minte' n-ar prea avea scop dacă mintea n-ar fi în stare să influenţeze într-un fel corpul fizic, sau să fie influenţată de el. Mai mult, dacă mintea este doar un 'epifenomen' — o trăsătură specifică, dar pasivă a stării fizice a creierului — adică un subprodus al corpului care nu poate avea nici un fel de influenţă asupra acestuia, atunci rolul său este lipsit de forţă şi până la urmă nesemnificativ. Pe de altă parte, dacă mintea ar fi în stare să influenţeze corpul în feluri care să producă ieşirea acestuia din constrângerile legilor fizicii,

însăşi acurateţea acestor legi fizice ar fi pusă la îndoială. Este deci dificil să susţinem descrierea drastic 'dualistă' după care mintea şi corpul ascultă două feluri diferite şi independente de legi. Chiar dacă legile fizice care guvernează acţiunile corpului permit o libertate în care mintea poate sa-i afecteze în mod relevant comportamentul, natura particulară a acestei libertăţi trebuie să fie ea însăşi un ingredient important al legilor fizice în discuţie. Orice este ceea ce controlează sau descrie mintea trebuie să fie o parte integrală a aceleiaşi mari scheme care guvernează, de asemenea, toate atributele materiale ale universului nostru.

Există cercetători[1] care pretind că, dacă ne referim la *minte* doar ca la un alt tip de 'substanţă' — cu toate că diferită de materie şi satisfăcând tipuri diferite de principii — comitem o 'eroare categorială' Ei ar putea atrage atenţia asupra unei analogii, după care corpul material ar trebui comparat cu un computer, iar mintea cu un program de computer. Într-adevăr, astfel de comparaţii pot fi utile acolo unde ele se potrivesc şi este desigur important să evităm confuziile între diferitele feluri de concepte. Cu toate acestea, a atrage atenţia asupra unei posibile 'erori categoriale' în cazul minţii şi al corpului nu soluţionează paradoxul.

Mai mult, există, în fizică, anumite concepte care pot fi într-adevăr echivalate între ele, deşi la prima vedere a face acest lucru pare o eroare categorială. Un exemplu de acest fel apare până şi în celebra ecuaţie a lui Einstein $E = mc^2$, care echivalează efectiv energia cu masa.Poate părea că o eroare categorială a apărut aici câtă vreme masa este măsura substanţei materiale actuale, pe când energia pare să fie o cantitate abstractă mai nebuloasă care descrie capacitatea unui corp de a efectua lucru mecanic. Şi totuşi, formula lui Einstein care le leagă este piatra unghiulară a fizicii moderne şi a fost confirmată experimental în numeroase tipuri de procese fizice. Un exemplu şi mai îndrăzneţ de potenţială eroare categorială în fizică apare odată cu conceptul de entropie (vezi de exemplu MN, capitolul 7). Entropia este definită într-un fel foarte subiectiv, fiind în esenţă o trăsătură a noţiunii de 'informaţie'; şi totuşi entropia se găseşte de asemenea legată de alte mărimi fizice mai 'materiale' în ecuaţii matematice determinate[2].

În mod asemănător, pare să nu existe nici un motiv să interzicem încercarea de a discuta noţiunea de 'minte' în funcţie de alte concepte fizice. Conştiinţa, în particular, pare să fie ceva ce este în mod obiectiv prezentă în asociaţie cu anumite obiecte fizice specifice — creiere umane vii în stare de veghe,

cel puţin — astfel că am putea anticipa un fel de descriere fizică a fenomenului, indiferent de cât suntem în prezent de departe de o înţelegere a lui. Unul dintre indiciile pe care le-am obţinut din discuţia din partea întâi a cărţii este că înţelegerea conştientă, în special, trebuie să conţină un fel de acţiune fizică non-algoritmică — cel puţin dacă e să urmăm concluziile în favoarea cărora am adus argumente şi anume ceva de tipul punctului de vedere $\mathcal{C}$ în loc de $\mathcal{A}$ $\mathcal{B}$ sau $\mathcal{D}$ (vezi 1.3). Va trebui să-i rog pe cititorii încă neconvinşi de argumentele mele să mă urmeze deocamdată în această explorare şi să vedem împreună care este teritoriul pe care acest punct de vedere ni-l oferă spre cercetare. Vom vedea că posibilităţile pe care acest drum ni le deschide nu sunt în nici un caz atât de nefavorabile pe cât ar aştepta unii şi că putem găsi multe lucruri în acest teritoriu care vor fi de mare interes în sine. Sper că, după ce vom întreprinde împreună această muncă de explorare, o să vă întoarceţi cu mai multă simpatie la argumentele — pe care le cred destul de puternice — pe care le-am prezentat până acum în această carte. Şi acum să pornim în explorare — folosind punctul de vedere $\mathcal{C}$ drept ghid!

4.2. Calculabilitate şi haos în fizica zilelor noastre

Precizia şi scopul legilor fizice, aşa cum sunt ele apreciate în prezent, este cu adevărat extraordinară. Dar ele nu conţin nici un fel de indicaţie a vreunei acţiuni care n-ar putea fi simulată computaţional. Cu toate acestea, noi va trebui să găsim printre posibilităţile pe care aceste legi ni le oferă o deschidere către acţiunile non-computaţionale ascunse care intervin în funcţionarea creierelor noastre. Voi amâna, pentru moment, o discuţie asupra naturii posibile a non-computabilităţii. Avem până acum toate motivele să credem că o astfel de discuţie va fi prea subtilă şi complicată pentru acest stadiu al drumului nostru. Mă voi întoarce la ea ceva mai târziu (paragrafele 7.9 şi 7.10). Ajunge deocamdată să spunem că pentru discuţia asupra non-computabilităţii minţii umane avem nevoie de o imagine esenţial diferită decât cea prezentă în teoriile fizice de care dispunem, fie ele clasice sau cuantice.

În fizica *clasică* avem posibilitatea ca în orice moment să determinăm toate datele necesare pentru definirea unui sistem fizic, iar evoluţia viitoare a acelui

sistem este nu numai complet determinată de aceste date, dar poate fi de asemenea *calculată* din ele prin metodele calculului Turing. Un astfel de calcul poate fi efectuat cel puţin în principiu, ţinând seama de două condiţii adiţionale.

Prima dintre aceste condiţii este ca datele să poată fi digitizate în mod adecvat, astfel încât parametrii continui ai teoriei să poată fi înlocuiţi, cu un grad de aproximare destul de bun, cu parametri discreţi (acest lucru se face în mod normal în simulările pe calculator ale sistemelor clasice).

A doua condiţie priveşte faptul că multe sisteme fizice sunt *haotice* — în sensul că, dacă vrem să calculăm comportamentul viitor cu un grad tolerabil de precizie, vom descoperi ca avem nevoie de date de un grad de precizie nerezonabil. Aşa cum am discutat anterior (vezi în special 1.7, de asemenea 3.10, 3.22), comportarea haotică într-un sistem cu funcţionare discretă *nu* furnizează tipul de 'non-calculabilitate' de care am avea nevoie.

Un sistem haotic (discret), chiar dacă dificil de calculat, este totuşi un sistem calculabil — după cum o arată faptul că astfel de sisteme sunt investigate în mod normal, în practică, cu ajutorul calculatoarelor electronice. Prima condiţie este deci legată de a doua; dacă gradul de precizie al aproximaţiei noastre discrete pentru parametrii continui ai teoriei ne satisface sau nu ca 'adecvare' este o decizie care depinde, într-un sistem haotic, de interesul nostru pentru comportarea lui *actuală* sau pentru comportarea lui *tipică*. Dacă cea din urmă ne este de ajuns — şi, aşa cum am argumentat în partea întâia, acest lucru pare să fie suficient pentru scopurile inteligenţei artificiale — atunci nu mai e nevoie să ne preocupe faptul că aproximaţiile noastre discrete nu sunt perfecte şi că erori mici în datele iniţiale pot conduce la erori foarte mari în funcţionarea ulterioară a sistemului. Dacă comportarea tipică a sistemului este *tot* ceea ce ne interesează, atunci cele două condiţii de mai sus par să ne asigure că nu avem în faţă nici o posibilitate serioasă privind existenţa non-calculabilităţii de tipul cerut în discuţia din partea întâia pentru orice sistem al fizicii clasice.

Pe de altă parte, n-ar trebui să neglijăm posibilitatea ca în comportarea haotică a unor sisteme matematice continue (luate ca model al unor sisteme fizice actuale) să existe ceva care să nu poată fi descris de nici o aproximaţie discretă. Nu ştiu ca un astfel de sistem să existe, dar chiar dacă ar exista, n-ar fi de ajutor punctului de vedere al inteligenţei artificiale — aşa cum acesta se prezintă în momentul de faţă — deoarece acesta depinde de modelarea prin calcul discret

(adică digital şi nu analogic, după cum am arătat în 1.8).

În fizica cuantică există de asemenea o anumită libertate de natură complet aleatoare, deasupra şi dincolo de comportarea deterministă (şi computaţională) pe care o descriu ecuaţiile teoriei cuantice (şi în principal ecuaţia lui Schrödinger). Tehnic, aceste ecuaţii nu sunt haotice, însă absenţa haosului este înlocuită de prezenţa componentelor aleatoare deja menţionate care se adaugă evoluţiei deterministe. Pe de altă parte, aşa cum am văzut, în particular în paragraful 3.18, astfel de ingrediente aleatorii nu par să încurajeze o acţiune nealgoritmică. Astfel, pare că nici fizica clasică nici cea cuantică, aşa cum sunt ele înţelese în prezent cel puţin, nu permit un spaţiu pentru comportamentul non-calculabil de tipul cerut, ceea ce ar însemna că ar trebui să căutăm în altă parte acţiunile non-computaţionale de care avem nevoie.

4.3. Conştiinţa: o nouă fizică sau un 'fenomen emergent'?

În partea întâi am argumentat (pe cazul particular al înţelegerii matematice) că fenomenul conştiinţei poate apărea doar în prezenţa unor procese fizice non-computaţionale care ar avea loc în creier. Pe de altă parte, trebuie să presupunem că astfel de procese, ca cele pe care le căutăm trebuie să aibă *de asemenea* loc în acţiunile materiei neînsufleţite, câtă vreme creierele umane vii sunt în ultimă instanţă concepute din aceeaşi materie şi satisfac aceleaşi legi fizice ca şi obiectele neînsufleţite din Univers. Trebuie deci să ne punem două întrebări. Mai întâi, cum se face că fenomenul conştiinţei pare să existe doar în creiere (sau în relaţie cu ele) — cu toate că nu trebuie să excludem posibilitatea unor alte sisteme fizice conştiente. În al doilea rând, trebuie să ne întrebăm cum se face că un asemenea ingredient important precum comportarea non-computaţională, aparent inerent — cel puţin potenţial — în acţiunea tuturor lucrurilor materiale, a scăpat până acum cu desăvârşire fizicienilor?

Răspunsul la prima întrebare va avea fără îndoială de-a face cu organizarea sofisticată şi complexă a creierului, însă acest lucru nu poate furniza o explicaţie de unul singur. După cum am sugerat şi până acum, voi considera ca organizarea creierului trebuie să fie potrivită (adecvată) pentru a folosi non-computabilitatea

Acest punct de vedere diferă esenţial de cel mai popular al inteligenţei artificiale tari[3] (punctul de vedere $\mathcal{A}$), după care conştiinţa conştientă ar fi un tip de 'fenomen emergent' care apare mai degrabă ca o trăsătură a unei complexităţi sau sofisticări suficiente şi nu presupune existenţa nici unui proces fizic adiacent diferit de cele 'normale' în materia neînsufleţită. Cazul prezentat în partea întâi argumentează în mod diferit, cerând să existe în creier o anumită organizare subtilă, alcătuită în aşa fel încât să poată funcţiona pe bazele deja menţionate ale fizicii non-computaţionale. Voi reveni cu detalii asupra acestui subiect în cele ce urmează (7.4 - 7.7).

În ce priveşte a doua întrebare, va trebui să ne aşteptăm la prezenţa unor vestigii de non- computabilitate în materia nevie. Cu toate acestea, fizica materiei aşa cum este ea astăzi, nu pare — la prima vedere cel puţin — să permită vreo comportare non-computaţională. În continuare voi încerca să explic pe larg cum anume o astfel de comportare ar putea fi compatibilă cu observaţiile de care dispunem în prezent. Pentru moment ne va fi util să discutăm un fenomen oarecum diferit însă analog din fizica deja cunoscută. Deşi fără conexiuni — sau, cel puţin, fără conexiuni *directe* — cu vreun tip de comportare non-computaţională, acest fenomen fizic seamănă foarte mult cu modelul nostru de non-coputabilitate prin aceea că, deşi prezent, este imposibil de pus în evidenţă în comportarea actuală a obiectelor obişnuite. Cu toate acestea el poate fi pus în evidenţă la un nivel corespunzător al observaţiilor noastre despre lume şi a modificat profund felul în care înţelegem acum universul. Povestea care urmează este, de fapt, esenţială pentru însăşi dezvoltarea ştiinţei moderne.

4.4. *Înclinarea* Einstein

Încă de pe timpul lui Isaac Newton, *gravitaţia* şi descrierea ei matematică minunat de precisă (elaborată pentru prima oară de către Newton în 1687) a jucat un rol cheie în dezvoltarea gândirii ştiinţifice. Odată ce această descriere matematică a fost stabilită, fenomenul gravitaţiei a servit drept model pentru descrierea altor procese fizice în care mişcarea corpurilor printr-un cadru spaţial fix (fără curbură) era văzută ca fiind controlată în mod precis de către forţe care acţionează asupra corpurilor, forţe de atracţie sau respingere mutuală între particule

individuale şi care guvernează mişcarea acestor particule în cele mai mici detalii. Ca rezultat al succesului extraordinar al teoriei gravitaţionale a lui Newton s-a ajuns la convingerea că *toate* procesele fizice pot fi descrise în acest fel; forţele electrice, magnetice, moleculare şi aşa mai departe, acţionând între particule şi controlându-le mişcările după acelaşi model — care funcţionase atât de frumos în cazul gravitaţiei.

În 1865 această imagine a fost drastic modificată, când marele fizician scoţian, James Clerk Maxwell a publicat un remarcabil set de ecuaţii care descriau comportarea câmpurilor electrice şi magnetice. Aceste câmpuri continue erau văzute acum ca având o existenţă separată, independentă de cea a diverselor particule discrete. Câmpul electromagnetic (aşa se numeşte combinaţia celor două câmpuri de care este vorba) este capabil să transporte energie prin spaţiul altfel gol sub formă de lumină, unde radio, raze X, etc. şi are acelaşi grad de realitate ca şi particulele newtoniene cu care este considerat coexistent. Totuşi, descrierea generală este cea a unor corpuri fizice (incluzând acum sub această etichetă şi câmpurile continue) care se mişcă într-un spaţiu fix sub influenţa interacţiunilor mutuale. În acest fel, imaginea generală impusă de modelul newtonian nu a fost substanţial alterată. Nici măcar teoria cuantică, introdusă în anii 1913-1926 de către Niels Bohr, Werner Heisenberg, Erwin Schrödinger, Paul Dirac şi alţii, nu a modificat radical acest fel de a privi lumea, cu toate că stranietăţile nu-i lipseau. Obiectele fizice au continuat să fie considerate drept lucruri materiale, acţionând unele asupra altora prin intermediul unor câmpuri de forţe, în timp ce toate acestea erau conţinute într-un acelaşi cadru spaţial fix, imobil şi neted.

În acelaşi timp cu inventarea mecanicii cuantice, Albert Einstein a fost cel care a reexaminat bazele teoriei newtoniene a gravitaţiei şi în final, în 1915 a publicat o nouă şi revoluţionară teorie care a introdus în fizică o imagine cu totul diferită. În teoria generală a relativităţii (vezi MN) gravitaţia nu mai este o forţă ci trebuie reprezentată ca un fel de *curbură* a spaţiului însuşi (de fapt a obiectului fizic numit spaţiu-timp) în care toate celelalte particule şi forţe coexistă.

Nu toţi fizicienii sunt mulţumiţi de această nouă imagine neliniştitoare. Mulţi au simţit că gravitaţia n-ar trebui tratată diferit de alte acţiuni fizice — mai ales că ea a slujit atâta vreme drept modelul prototip al interacţiunilor, după care toate celelalte teorii au fost modelate. O altă problemă este că gravitaţia este deosebit de slabă în comparaţie cu alte forţe: de exemplu, forţa gravitaţională dintre

electron şi proton într-un atom de hidrogen este mai slabă decât forţa electrică dintre cele două particule cu un factor de aproximativ:

1/28 500 000 000 000 000 000 000 000 000 000 000 000 000

Astfel, gravitaţia este pur şi simplu imposibil de observat la nivelul particulelor individuale care constituie materia!

O întrebare care poate fi pusă uneori este dacă nu cumva gravitaţia ar fi un anumit fel de efect rezidual, rezultând probabil din anularea aproape completă a celorlalte forţe implicate (anumite forţe de acest fel sunt cunoscute; de exemplu forţa Van der Waals, punţile de hidrogen, forţa London). Dacă lucrurile ar sta aşa, atunci departe de a fi un fenomen fizic diferit de oricare altul şi care ar trebui descris matematic într-un mod cu totul special gravitaţia n-ar avea o existenţă individualizată, ci ar fi mai curând ceva de tipul fenomenelor emergente. Un exemplu al acestui punct de vedere este teoria unui mare savant şi luptător pentru drepturile omului, Andrei Sakharov.

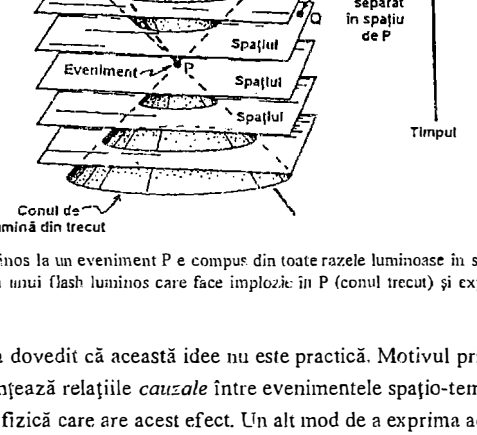


Fig. 4.1. Conul luminos la un eveniment P e compus din toate razele luminoase în spaţiu-timp prin P. Reprezintă istoria unui flash luminos care face implozie în P (conul trecut) şi explodează din nou (conul viitor).

Totuşi, s-a dovedit că această idee nu este practică. Motivul principal este că gravitaţia influenţează relaţiile *cauzale* între evenimentele spaţio-temporale şi este *singura* mărime fizică care are acest efect. Un alt mod de a exprima acest lucru este să spunem că gravitaţia are capacitatea unică de a înclina (răsuci) conurile luminoase (vom vedea în curând ce înseamnă asta). Nici un alt câmp fizic,

altul decât gravitaţia şi nici vreo combinaţie oarecare de câmpuri non-gravitaţionale nu poate înclina conurile luminoase.

Dar ce înseamnă această 'înclinare[1] a conurilor luminoase'? Ce sunt „relaţiile cauzale între evenimentele spaţio-temporale"? Pentru a explica toate acestea avem nevoie de o mică digresiune, de mare importanţă în cele ce urmează. Unii dintre cititori sunt probabil familiarizaţi cu aceste concepte. Voi da, în cele ce urmează, doar o scurtă explicaţie pentru a prezenta celorlalţi ideile necesare (pentru o discuţie mai completă se poate vedea capitolul 5 din MN). În figura 4.1 am desenat, într-o diagramă spaţiu-timp, un con de lumină obişnuit. Timpul este reprezentat pe diagramă ca o progresie de momente din josul paginii spre partea de sus, iar spaţiul este reprezentat ca întinzându-se orizontal. Un punct în această diagramă spaţiu-timp reprezintă un *eveniment*, adică un punct spaţial determinat la un anumit moment. Evenimentele au deci durată temporală zero şi extensie spaţială nulă. Conul luminos, centrat în evenimentul *P* reprezintă istoria spaţio-temporală a unui puls sferic de lumină care se află în implozie în punctul *P*, sau explodează din punctul *P*, în ambele cazuri, desigur, cu viteza luminii. Deci, întregul con luminos pornit din *P* este constituit din acele raze de lumină care întâlnesc evenimentul *P* în istoriile lor individuale.

Conul de lumină *P* are două părţi, conul de lumină *trecut* (în diagramele din MN această parte nu a fost reprezentată) reprezentând semnalul luminos în implozie şi conul de lumină *viitor*, reprezentând explozia. Conform teoriei relativităţii, evenimentele care pot avea o influenţă cauzală asupra evenimentului *P* sunt cele care se găsesc în interiorul conului luminos viitor sau pe suprafaţa acestuia. Acele evenimente care se găsesc în afara ambelor conuri sunt cele care nu pot nici influenţa evenimentul *P*, nici nu pot fi influenţate de el. Aceste evenimente se numesc separate spaţial.

Trebuie să subliniem că aceste noţiuni de cauzalitate sunt trăsături ale teoriei relativităţii şi nu sunt pertinente în fizica newtoniană. În cadrul elaborat de Newton nu există viteză de transfer limitată a informaţiei. Doar teoria relativităţii stabileşte o astfel de viteză limită, viteza luminii. Este un principiu fundamental al teoriei relativităţii faptul că nici un efect cauzal nu se poate propaga cu o viteză mai

[1] Propunem termenul de *înclinare* sau *deformare* Einstein pentru noţiunea „Einstein *tilt*" care reprezintă încă un concept 'nou' în literatura românească de specialitate (n.t.)

mare decât viteza luminii.

Pe de altă parte, trebuie să fim atenţi ce anume desemnăm prin 'viteza luminii' Semnalele luminoase actuale sunt uşor încetinite atunci când trec printr-un mediu de refracţie, ca de exemplu sticla. Într-un astfel de mediu viteza semnalului luminos va fi mai mică decât ce numim aici 'viteza luminii' şi este posibil pentru un corp fizic sau pentru un semnal fizic diferit de semnalul luminos să depăşească limita cu care se propagă lumina într-un astfel de mediu. Un astfel de fenomen poate fi observat în anumite experimente şi este numit radiaţie Cerenkov. În acest caz particulele se propagă într-un mediu de refracţie cu o viteză ceva mai mică decât viteza luminii în vid, dar mai mare decât viteza cu care lumina se propagă prin acel mediu. Undele de şoc luminoase care apar sunt radiaţia Cerenkov.

Pentru a evita confuzia mă voi referi la viteza luminii atunci când vorbesc despre valoarea ei absolută (în vid). Conurile luminoase determină viteza absolută, dar nu determină în mod necesar şi viteza actuală a luminii. Într-un mediu oarecare, viteza actuală a luminii este ceva mai mică decât cea absolută şi este de asemenea mai mică decât cea a particulelor accelerate pentru a produce radiaţia Cerenkov Viteza absolută este cea care fixează limita pentru toate semnalele sau corpurile materiale şi cu toate că lumina nu are în toate mediile această viteză, aceasta este viteza cu care lumina călătoreşte în spaţiul vid.

Teoria la care ne referim aici este de fapt teoria relativităţii restrânse, cea în care gravitaţia este absentă. Conurile luminoase în relativitatea restrânsă sunt cu toate aranjate uniform, aşa cum este apar în figura 4.2, iar spaţiu — timpul se numeşte spaţiu Minkowski. În teoria relativităţii generalizate discuţia precedentă este încă valabilă atâta vreme cât ne referim la viteza absolută ca fiind determinată de către situaţia spaţio-temporală a conurilor de lumină.

În prezenţa gravitaţiei conurile pot deveni neuniforme în distribuţia lor, aşa cum este reprezentat în figura 4.3. La acest fenomen m-am referit când am vorbit de înclinarea conurilor luminoase.

Un mod destul de răspândit de a gândi această înclinare a conurilor luminoase este următorul: să spunem că viteza luminii — sau o viteză absolută — variază dintr-un loc într-altul. Această viteză depinde de asemenea de direcţia mişcării. În acest fel am putea acum să ne gândim la *viteza absolută* ca la ceva analog 'vitezei actuale a luminii' la care ne-am referit atunci când am vorbit de

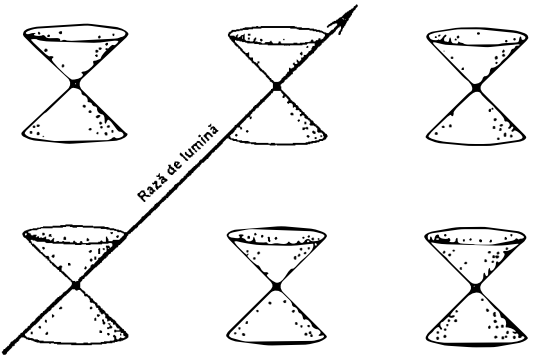


Fig. 4.2. Spaţiul Minkowski: spaţiu-timpul relativităţii speciale. Conurile de lumină sunt toate aranjate uniform.

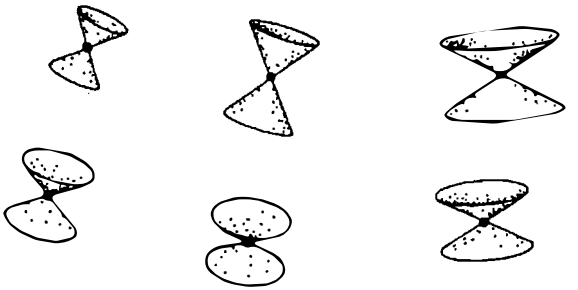


Fig. 4.3. Conurile luminoase răsucite ale relativităţii generalizate

trecerea luminii printr-un mediu de refracţie. Astfel, am putea considera câmpul gravitaţional drept un fel de mediu de refracţie atot-cuprinzător care afectează nu numai comportarea luminii, dar şi a tuturor particulelor materiale şi semnalelor (Este remarcabil că Newton însuşi a sugerat o idee de această natură. Vezi Întrebările 18-22 de la sfârşitul cărţii a treia în Optica din 1730). Într-adevăr, o descriere de acest tip a efectelor gravitaţiei a fost deseori oferită. Ea funcţionează în anumite aspecte, însă nu este întru totul satisfăcătoare şi în anumite privinţe poate oferi o imagine deformată şi înşelătoare a relativităţii generalizate.

În primul rând, cu toate că acest 'mediu gravitaţional de refracţie' poate

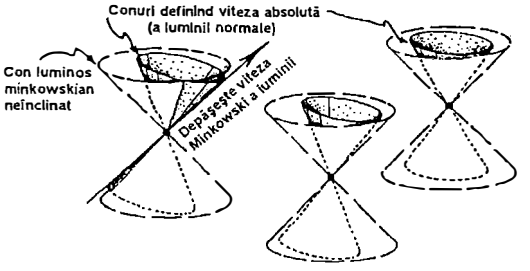


Fig. 4.4. Propagarea luminii în teoria relativităţii generalizate poate fi considerată ca un efect al unui 'mediu de refracţie', în cadrul spaţiului Minkowski, fără a viola legea privind viteza...

fi considerat deseori ca provocând o încetinire a vitezei absolute, aşa cum este cazul unui mediu de refracţie obişnuit, există exemple importante (câmpul gravitaţional al unei mase izolate la distanţe mari) în care descrierea nu funcţionează. 'Mediul de refracţie' trebuie, în asemenea cazuri să fie în stare să *mărească* viteza absolută în anumite locuri (Penrose 1989; vezi figura 4.4). Acest lucru nu este ceva care să poată fi obţinut în cadrul relativităţii restrânse în care un mediu de refracţie, oricât de exotic, nu va putea niciodată să accelereze semnale la viteze mai mari decât viteza luminii în vid fără a viola principiile de cauzalitate. O astfel de acceleraţie ar permite semnalelor să se propage în afara conurilor de lumină Minkowski (care desemnează mişcarea în afara oricărui mediu de refracţie), ceea ce nu este permis. În particular, efectul gravitaţional de deformare a conurilor luminoase descris mai sus nu poate fi interpretat ca un rezultat rezidual al unor câmpuri non-gravitaţionale.

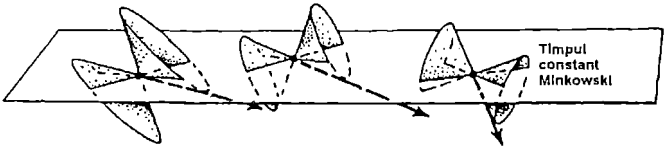


Fig. 4.5. În principiu, înclinarea conurilor luminoase poate fi atât de pronunţată încât semnalele luminoase să poată fi propagate în trecutul Minkowski.

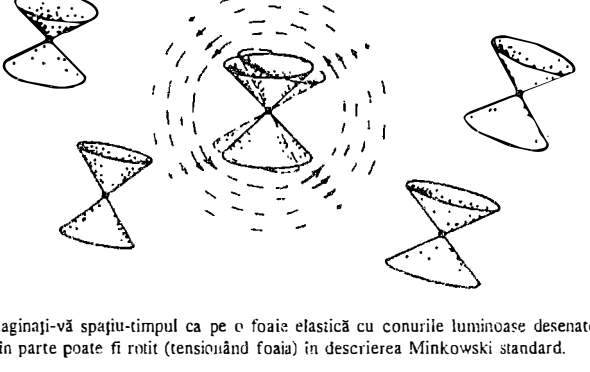


Fig. 4.6. Imaginaţi-vă spaţiu-timpul ca pe o foaie elastică cu conurile luminoase desenate pe ea. Fiecare con în parte poate fi rotit (tensionând foaia) în descrierea Minkowski standard.

Există şi alte numeroase situaţii extreme în care nu este posibil să descriem deformarea conurilor luminoase în acest fel, chiar dacă am permite vitezei absolute să fie 'accelerată' în anumite direcţii. În figura 4.5 este ilustrată o imposibilitate de acest fel, conurile de lumină fiind deformate într-un mod absurd.

Este cazul unor situaţii în care apar 'violări de cauzalitate' — astfel apare, de exemplu posibilitatea teoretică ca observatorul să trimită semnale în propriul său trecut (vezi şi figura 7.15 în capitolul 7). În mod cu totul remarcabil, consideraţii de această natură vor avea relevanţă în discuţiile noastre viitoare, de exemplu în 7.10.

Un punct interesant şi mai subtil al discuţiei este că 'gradul de înclinare' al unui con luminos individual nu este ceva măsurabil şi deci nu are nici un sens fizic să-l tratăm ca pe o încetinire sau accelerare actuală a vitezei absolute. Acest lucru este ilustrat cel mai bine dacă ne gândim la figura 4.3 ca la o imagine desenată pe o foaie elastică. Orice con luminos individual poate fi rotit sau deformat în vecinătatea vârfului propriu (vezi figura 4.6) până când poziţia lui redevine 'verticală', ca în imaginile normale ale spaţiului Minkowski în relativitatea restrânsă (figura 4.2). Nu există nici o metodă de a spune cu ajutorul experimentelor locale dacă conul luminos al oricărui eveniment este 'înclinat' sau nu. Dacă ne-am gândi la efectul de înclinare ca datorat cu adevărat unui 'mediu gravitaţional' ar trebui să explicăm de ce acest mediu produce efectul curios de a fi neobservabil în orice punct-eveniment spaţio-temporal. În particular, chiar şi situaţiile extreme ilustrate în figura 4.5, pentru care ideea mediului gravitaţional

nu funcţionează, nu sunt — dacă considerăm doar un singur con luminos — fizic diferite de ceea ce se întâmplă într-o situaţie ca cea din spaţiul Minkowski în care conurile luminoase nu sunt înclinate deloc.

În general, totuşi, putem roti un con de lumină particular la orientarea Minkowski doar cu preţul de a avea câteva dintre conurile învecinate îndepărtate de la orientarea lor Minkowski. Există în general o constrângere matematică care face imposibilă deformarea foii elastice în aşa fel încât să aducă toate conurile luminoase în aranjamentul Minkowski standard arătat în figura 4.2. Pentru spaţiu-timpul patrudimensional aceasta constrângere este descrisă de un obiect matematic numit *tensorul Weyl conform* (pentru care în **MN** a fost folosită notaţia *WEYL*).(Tensorul Weyl descrie doar jumătate de informaţie — jumătatea 'conformă'- care este conţinută în tensorul Riemann care descrie curbura spaţiu-timpului. Toţi aceşti termeni însă nu trebuie să vă îngrijoreze. Deocamdată nu avem nevoie de ei). Doar dacă *WEYL* este zero putem să rotim *toate* conurile luminoase într-un aranjament Minkowski. Tensorul *WEYL* măsoară câmpul gravitaţional — şi deci chiar *câmpul gravitaţional* este cel care ne furnizează constrângerea şi ne împiedică să scăpam de efectul de 'înclinare' a conurilor luminoase.

Această cantitate tensorială este cu siguranţă măsurabilă fizic. Câmpul gravitaţional *WEYL* al Lunii, de exemplu, îşi exercită distorsiunea asupra Pământului — dând principala contribuţie la fenomenul mareelor (vezi figura 5.5 din **MN**). Acest efect nu este legat direct de Înclinarea conurilor luminoase; ci este o trăsătură a efectelor newtoniene ale gravitaţiei. Mai relevant este un alt efect observaţional, cel de lentilă gravitaţională (*gravitational lensing*), o trăsătură caracteristică a teoriei lui Einstein. Primul exemplu de acest fel a fost văzut de expediţia condusă de Arthur Eddington în 1919, care măsura distorsiunea produsă în câmpul stelelor de către câmpul gravitaţional al Soarelui. Distorsiunea locală a acestui câmp cadru este de aşa natură încât o formă circulară mică în câmpul actual va fi distorsionată într-una eliptică (vezi figura 4.7). Este o observaţie aproape directă a efectelor lui *WEYL* asupra structurii conurilor de lumină ale spaţiu-timpului. În ultima vreme,efectul de lentilă gravitaţională a devenit o unealtă deosebit de importantă în astronomia observaţională şi în cosmologie. Lumina de la un quasar îndepărtat este uneori distorsionată de prezenţa unei galaxii interpuse pe traseu (figura 4.8). Observaţiile distorsionate ale formei quasarului, împreună cu efectele de întârziere ne pot da informaţii interesante asupra distanţelor, maselor

etc. Toate acestea furnizează suficiente dovezi experimentale pentru înclinarea conurilor luminoase ca fenomen actual, ca şi pentru efectele măsurabile ale tensorului *WEYL*.

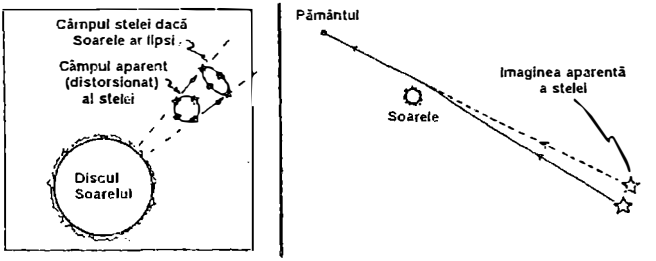


Fig. 4.7. Un efect observaţional direct al înclinării conurilor luminoase. Tensorul de curbură al spaţiu-timpului WEYL intervine ca distorsiune a câmpului stelei depărtate.

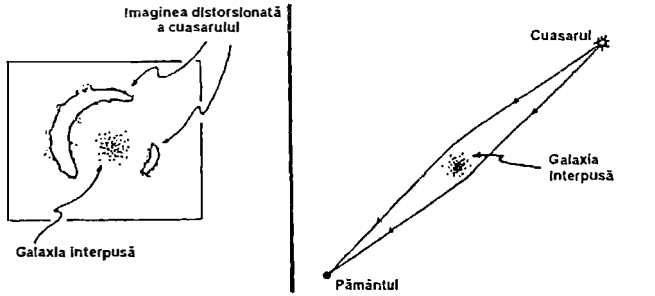


Fig. 4.8. Efectul de curbare a razei de lumină este foarte important în astronomia observaţională. Masa unei galaxii interpuse poate fi dedusă din distorsiunea imaginii unui quasar îndepărtat.

Observaţiile care urmează ilustrează faptul că înclinarea conurilor luminoase, adică distorsiunea cauzală datorată gravitaţiei, este nu numai un fenomen subtil ci unul *real*; ea nu poate fi eliminată printr-o explicaţie de tipul proprietăţii 'emergente' (care ar apărea dacă conglomeratele de materie ar fi destul de largi). Gravitaţia are deci un caracter specific şi unic printre procesele fizice. Chiar dacă nu este direct discernabilă la nivelul forţelor care sunt importante pentru particulele elementare, ea este permanent acolo. Nimic altceva din ce cunoaştem

astăzi în fizică în afara gravitaţiei nu ar putea înclina conurile luminoase, deci gravitaţia este ceva total *diferit* de toate celelalte forţe cunoscute sau influenţe fizice. Conform teoriei relativităţii generalizate, efectul de înclinare al conurilor luminoase trebuie să existe în preajma oricărei cantităţi de materie, chiar în cel mai mic fir de nisip. Chiar electronii individuali vor afecta conurile luminoase. Însă efectul în astfel de obiecte este atât de mic încât nu poate avea nici un fel de efect observabil.

Efectele gravitaţiei au fost observate în obiecte considerabil mai mari decât firele de nisip dar totuşi mai mici decât Luna. Într-un experiment celebru efectuat în 1798, Henry Cavendish a măsurat atracţia gravitaţională a unei sfere de aproximativ 10^5 grame (experimentul său se baza de fapt pe unul precedent imaginat de John Michell). Cu tehnologia modernă este însă posibil să determinăm efecte gravitaţionale ale unor mase mult mai mici (vezi, de exemplu, Cooke, 1988). Totuşi, orice detectare a efectului de înclinare a conurilor luminoase în oricare din aceste situaţii este deocamdată imposibilă pentru tehnicile actuale. Numai în cazul maselor cu mult mai mari acest efect poate fi observat; cu toate acestea, prezenţa sa în orice corp material este o implicaţie clară a teoriei einsteiniene.

Efectele de detaliu ale gravitaţiei nu pot fi simulate de nici o combinaţie de alte câmpuri. Ele au un caracter unic şi nu există nici o modalitate de a le privi ca pe un fenomen emergent sau secundar rezultat din vreun alt proces fizic mai important. Efectele gravitaţiei sunt descrise chiar de structura spaţiu-timpului care era înainte văzută drept cadrul fix pentru orice activitate fizică. În universul newtonian gravitaţia nu avea nimic special — chiar dacă oferea modelul pentru toate celelalte forţe. În universul lui Einstein (care este imaginea confirmată observaţional de către fizica contemporană) gravitaţia este privită ca ceva cu totul diferit: nici pe departe un fenomen emergent, ci unul cu un caracter specific determinat.

În ciuda faptului că gravitaţia este diferită de toate celelalte forţe fizice, există o armonie profundă care o integrează celorlalte interacţiuni cunoscute. Teoria lui Einstein nu este ceva complet diferit de celelalte legi ale fizicii, ci doar le prezintă pe acestea într-o lumină cu totul diferită (lucru valabil mai ales pentru legile de conservare a energiei, impulsului şi momentului cinetic).

Posibilitatea integrării gravitaţiei einsteiniene cu restul fizicii explică într-o oarecare măsură ironia faptului demonstrat tot de Einstein: deşi gravitaţia

newtoniană a furnizat *paradigma* pentru restul fizicii, ea este de fapt diferită de restul fizicii! Mai presus de toate, lecţia lui Einstein ne-a învăţat că nu este bine să ajungem să credem — în nici un stadiu al înţelegerii noastre — că am găsit deja imaginea fizică corectă sau adecvată.

Putem oare spera să avem ceva asemănător de învăţat în ceea ce priveşte fenomenul conştiinţei? Dacă da, atunci ar fi că nu de *masă* avem nevoie pentru ca fenomenul să devină aparent — sau nu numai de masă — ci de un anumit fel de organizare fizică subtilă. Luând în considerare argumentele din partea întâia, o astfel de organizare ar trebui să conţină sau să folosească un element non-computaţional deja prezent în comportarea obişnuită a materiei — un ingredient care, asemenea înclinării conurilor luminoase în relativitatea generalizată ar fi scăpat cu desăvârşire atenţiei dacă această atenţie ar fi fost îndreptată doar spre comportarea particulelor elementare.

Dar are vreo legătură înclinarea conurilor luminoase cu non-computabilitatea? Vom explora un aspect interesant al acestei probleme în paragraful 7.10; deocamdată nu putem spune decât că nu există nici o legătură, cu excepţia faptului că putem să găsim o morală: ne întâlnim uneori în fizică cu o proprietate fundamentală cu desăvârşire nouă, complet diferită de oricare alta din cele cunoscute, ascunsă până nu demult în comportamentul familiar al materiei.

Numeroase consideraţii matematice sofisticate şi argumente fizice subtile l-au ajutat pe Einstein să ajungă la acest punct de vedere revoluţionar — cel mai important argument însă era bine cunoscut (deşi neapreciat) încă de pe vremea lui Galileo Galilei (principiul echivalenţei: toate corpurile cad cu aceeaşi viteză în câmpul gravitaţional). În plus, condiţia necesară pentru succesul ideilor lui Einstein a fost că ele trebuiau să fie compatibile cu tot ceea ce se ştia în fizica timpului său.

Într-un mod analog, putem considera că există o acţiune non-computaţională ascunsă undeva în comportamentul lucrurilor. Pentru ca speculaţia noastră să aibă speranţe de succes trebuie de asemenea să fie motivată de consideraţii puternice, foarte probabil de o considerabilă sofisticare matematică sau subtilitate fizică şi va trebui de asemenea să fie consistentă cu fenomenele fizice cunoscute astăzi. Vom vedea în cele ce urmează cât de departe putem merge pe drumul către o asemenea teorie.

Mai întâi, ca o precondiţie, să aruncăm o privire asupra atracţiei enorme pe care ideile computaţionale o au asupra fizicii contemporane. Nu fără ironie vom

vedea că relativitatea generalizată reprezintă unul dintre exemplele cele mai bune pe care Natura ni le pune la dispoziţie.

4.5. Calcul şi fizică

La aproximativ 30 000 de ani lumină de Pământ, în constelaţia Acvilei, două stele moarte de o densitate incredibilă orbitează una în jurul alteia. Materia lor este atât de densă încât o minge de tenis din această substanţă ar avea o masă comparabilă cu cea a satelitului lui Marte, Deimos. Cele două stele — numite stele neutronice — se rotesc una în jurul alteia în aproximativ 7 ore, 45 de minute şi 6,981 6132 secunde, iar masele lor sunt de 1,44 1 şi, respectiv, 1,3874 ori masa Soarelui (cu o eroare posibilă de cel mult 7 unităţi în ultima zecimală). Prima dintre aceste stele emite un puls de radiaţie electromagnetică (unde radio), în direcţia noastră, la fiecare 59 de milisecunde, ceea ce indică faptul că se roteşte în jurul axei sale de 17 ori într-o secundă. Acest obiect este ceeea ce se numeşte un pulsar, iar sistemul binar este cunoscut sub numele (celebru) de sistemul pulsar binar PSR 1913+16.

Pulsarii sunt cunoscuţi din 1967, când au fost descoperiţi de Joselyn Bell şi Anthony Hewish de la observatorul radio din Cambridge. Sunt cu adevărat nişte obiecte remarcabile! Stelele neutronice rezultă în mod normal din colapsul gravitaţional al nucleului unei stele gigante roşii, care poate produce o violentă explozie de supernovă. Densitatea unor astfel de stele este incredibilă, materia lor constând din particule nucleare, în special neutroni, astfel încât densitatea totală a unei stele neutronice este de acelaşi ordin de mărime cu densitatea neutronului însuşi. În timpul colapsului, liniile de flux ale câmpului magnetic sunt 'capturate' în substanţa stelei, astfel încât, datorită compresiei enorme care apare în steaua neutronică rezultantă, câmpul magnetic este extraordinar de puternic. Liniile câmpului magnetic vor emerge din polul nord al stelei şi, după ce vor străbate distanţe considerabile în spaţiu, se vor întoarce în polul magnetic sud (vezi figura 4.9).

Colapsul stelei va avea ca rezultat faptul că viteza sa de rotaţie va fi de asemenea mult mai mare (un efect al conservării momentului cinetic). În cazul pulsarului la care ne-am referit mai înainte (care are aproximativ 20 de kilometri

în diametru), această viteză este de cca 17 rotaţii pe secundă. Ca rezultat, câmpul magnetic extrem de puternic al pulsarului este răsucit împrejur de 17 ori pe secundă în timp ce liniile de câmp din interiorul stelei rămân fixate de corpul acesteia. În afara stelei, liniile de câmp duc particulele încărcate cu ele de jur împrejur. La o anumită distanţă de corpul stelei viteza cu care aceste particule se mişcă trebuie să se apropie de viteza luminii. Când acest lucru se întâmplă, particulele încărcate încep să radieze violent, iar undele radio puternice pe care ele le emit sunt transmise, ca dintr-un far gigantic, la distanţe enorme. Unele dintre semnale ating Pamântul şi sunt observate de astronomi ca succesiune de 'clicuri' radio, caracteristice unui pulsar (figura 4.10)

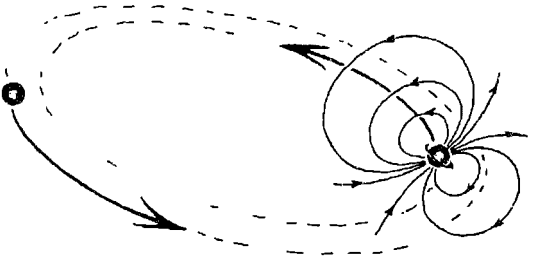


Fig. 4.9. PSR 1913+16. Două stele neutronice orbitează una în jurul celeilalte, una fiind un pulsar cu un câmp magnetic imens care prinde particulele din jur.

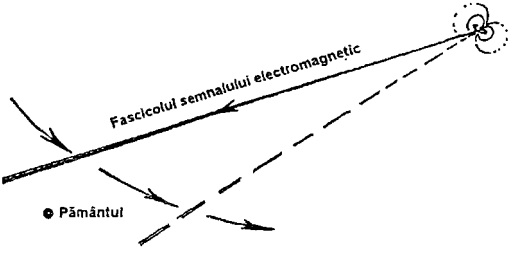


Fig. 4.10. Particulele încărcate capturate se rotesc de jur împrejurul pulsarului şi emit un semnal electromagnetic a cărui raze mătură Pamântul cu frecvenţa de 17 Hz.

Vitezele de rotaţie ale pulsarilor sunt extrem de stabile; ele ne furnizează

ceasuri ale căror precizie ajunge din urmă sau chiar depăşeşte precizia celor mai bune ceasuri nucleare care au fost construite până acum. (Un bun ceas pulsar poate grăbi sau întârzia mai puţin de 10^{-12} secunde într-un an.) Dacă pulsarul se întâmplă să fie parte a unui sistem binar, cum e cazul cu PSR 1913+16, atunci mişcarea sa pe orbită împrejurul companionului său poate fi monitorizată îndeaproape folosind efectul Doppler; rata 'ciicurilor' înregistrate este puţin mai mare atunci când pulsarul se apropie de noi decât atunci când se îndepărtează.

În cazul lui PSR 1913+16 a fost posibil să descriem cu o precizie remarcabilă cele două orbite ale stelelor şi să verificăm observaţional o serie de predicţii ale teoriei relativităţii generalizate. Printre acestea, un efect cunoscut sub numele de 'deplasarea periheliului' — observat la sfârşitul secolului trecut ca o comportare anormală a planetei Mercur în mişcarea ei împrejurul Soarelui (şi explicat de Einstein în 1916, când a constituit de altfel şi primul test al teoriei sale. De asemenea, se mai pot observa diferite tipuri de 'devieri' relativiste care afectează axele de rotaţie s.a.m.d. Teoria lui Einstein ne dă o imagine foarte clară (deterministă şi calculabilă) a felului în care două corpuri mici ar trebui să comporte atunci când se mişcă pe orbite confocale. Este posibil să calculăm această mişcare cu un grad înalt de acurateţe, folosind metode sofisticate de aproximaţie şi diferite tehnici de calcul standard. În calculul nostru vor exista câţiva parametri necunoscuţi, masele şi mişcările iniţiale ale stelelor, însă avem destule date din semnalele pe care pulsarul le emite pentru ca aceste mărimi să fie fizate cu precizie foarte bună. Potrivirea remarcabilă între descrierea rezultată în urma calculelor şi informaţiile detaliate primite sub forma semnalelor recepţionate de la pulsar ne furnizează un suport solid în favoarea teoriei relativităţii generalizate.

Există un alt efect al relativităţii generalizare la care nu m-am referit încă şi care joacă un rol important în dinamica pulsarului binar: *radiaţia gravitaţională*. Am subliniat în secţiunea anterioară cum diferă gravitaţia de toate celelalte câmpuri fizice. Putem găsi însă şi aspecte în care gravitaţia şi electromagnetismul sunt similare. O proprietate importantă a câmpurilor electromagnetice este că ele pot exista în formă ondulatorie, propagându-se prin spaţiu ca unde radio. După teoria clasică a lui Maxwell, astfel de unde vor emana din orice sistem de particule încărcate orbitând mutual, care interacţionează una cu alta prin forţe electromagnetice. În mod similar, teoria relativităţii generalizate prezice că orice

sistem de corpuri care gravitează pe orbite concentrice şi între care există doar interacţiuni gravitaţionale va emite unde gravitaţionale. În mod normal, acestea unde ar fi extrem de slabe. Cea mai puternică sursă de radiaţie gravitaţională în sistemul solar apare din mişcarea planetei Jupiter în jurul Soarelui, însă cantitatea de energie pe care sistemul Jupiter — Soare o emite în această formă este asemănătoare cu cea a unui bec electric de 40 de waţi!

Totuşi, în alte circumstanţe, ca de exemplu în cazul lui PSR 1913+16, situaţia este foarte diferită, iar radiaţia gravitaţională are, de această dată, un rol semnificativ Aici, teoria lui Einstein ne dă o bună predicţie a naturii radiaţiei gravitaţionale care va fi emisă de către sistem şi a energiei transportate. Pierderea de energie trebuie să rezulte dintr-o mişcare de spiralare înspre interior a celor două stele neutronice şi o accelerare corespunzătoare a perioadei de roataţie. Joseph Taylor şi Russell Hulse au observat pentru prima dată acest pulsar binar cu imensul telescop radio Aricebo din Puerto Rico în 1974. De atunci, perioada de rotaţie a fost atent monitorizată de Taylor şi colegii săi, iar acceleraţia este în acord foarte bun cu predicţiile relativităţii generalizate (vezi figura 4.11). Pentru acest lucru, Hulse şi Taylor au luat în 1993 Premiul Nobel pentru fizică. De fapt, cu trecerea timpului, acumulările de date au dus la o confirmare din ce în ce mai puternică a teoriei lui Einstein. Într-adevăr, dacă considerăm acum sistemul ca întreg şi-l comparăm cu comportamentul calculat pe baza teoriei lui Einstein ca întreg — de la aspectele newtoniene ale orbitelor, prin corecţiile acestor orbite cu efectele relativiste standard şi până la efectele asupra orbitelor datorate pierderii de energie în radiaţie gravitaţională — vom găsi că teoria este confirmată până la o eroare care nu este mai mare de 10^{-14} Astfel, teoria lui Einstein se dovedeşte a fi, în acest sens, cel puţin, cea mai exact testată teorie ştiinţifică!

În acest exemplu avem un sistem remarcabil de 'curat', în care avem nevoie doar de relativitatea generalizată pentru a efectua calculele. Complicaţiile care ar putea apărea din constituţia internă a corpurilor, câmpurilor magnetice sau unui mediu suplimentar, nu au vreun efect semnificativ asupra mişcărilor. Mai mult, doar două corpuri şi câmpurile lor gravitaţionale sunt implicate în problema noastră, ceea ce ne permite să calculăm direct comportarea aşteptată în toate detaliile. Până la urmă, ne aflăm în faţa celui mai bun exemplu de acord între un model teoretic calculat şi comportarea observată experimental — în cazul unui număr mic de corpuri.

Când numărul de corpuri într-un sistem fizic este considerabil mai mare decât acesta, putem încă, cu ajutorul calculatoarelor moderne, modela comportamentele sistemelor într-un mod oarecum detaliat. În particular, mişcarea planetelor din sistemul solar, împreună cu sateliţii lor mai senificativi a fost modelată printr-un calcul foarte detaliat de Irwin Shapiro şi colegii săi. Acesta s-a dovedit un alt test important al relativităţii generalizate. Şi, din nou, teoria lui Einstein s-a dovedit în acord cu datele observaţionale şi a explicat unele mici deviaţii, de la comportamentul observat, care dădeau de furcă descrierii newtoniene.

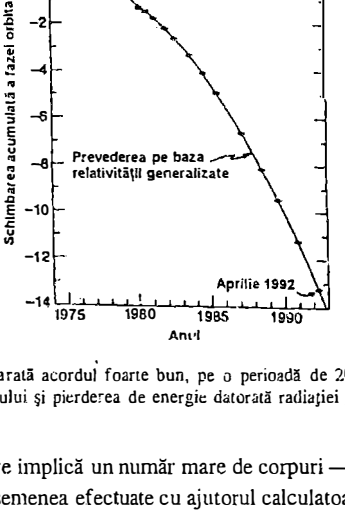


Fig. 4.11. Acest grafic arată acordul foarte bun, pe o perioadă de 20 de ani, dintre accelerarea mişcării orbitale a pulsarului şi pierderea de energie datorată radiaţiei gravitaţionale

Calculele care implică un număr mare de corpuri — uneori de ordinul unui milion — pot fi de asemenea efectuate cu ajutorul calculatoarelor moderne. Ele vor fi însă bazate mai ales — deşi nu exclusiv — pe teoria newtoniană. Trebuie făcute anumite presupoziţii simplificatoare în legătură cu felul în care ar trebui să aproximăm efectele mai multor particule printr-un fel de mediere, în loc să trebuiască să calculăm efectele fiecărei particule în parte asupra fiecărei alte particule. Astfel de calcule sunt un lucru comun în astrofizică, fie că este vorba despre formarea detaliată a stelelor şi galaxiilor, fie de condensarea materiei în

Universul timpuriu, înainte de formarea galaxiilor.

Trebuie însă să observăm o mare diferență între ceea ce aceste calcule încearcă să obțină. În cazurile de față vom fi mai puțin interesați de comportamentul *actual* a vreunui sistem cât de o evoluție *tipică*. Ca și în cazul considerațiilor noastre anterioare privind sistemele haotice, o comportare tipică este cel mai bun lucru pe care-l putem obține. Vom putea prin astfel de metode să testăm diferite ipoteze științifice privind constituirea și distribuția inițială a materiei în Univers, văzând cât de bine se potrivește modelarea lor cu ceea ce se poate observa astăzi. Nimeni nu așteaptă într-un astfel de caz o potrivire detaliată, ci doar aparențe generale asemănătoare și diferiți parametri statistici care pot fi comparați între model și observații.

Situația extremă de acest tip apare atunci când numărul particulelor este atât de mare încât tratarea lor individuală este fără speranță și ansamblul lor trebuie modelat statistic. Tratamentul matematic obișnuit al unui gaz, de pildă, consideră *ansambluri statistice* ale diferitelor colecții de mișcări de particule. Mărimile fizice ca temperatura, presiunea, entropia etc., sunt proprietăți ale unor astfel de ansambluri. Ele pot fi tratate ca parte a unui sistem calculabil, unde proprietățile evoluționare ale ansamblului sunt descrise statistic.

Pe lângă ecuațiile dinamice relevante (Newton, Maxwell, Einstein etc.) există, în astfel de circumstanțe, un alt principiu fizic care trebuie implicat. Este vorba despre *a doua lege a termodinamicii*[4]. Această lege elimină stările inițiale ale mișcărilor particulelor care ar duce la evoluții viitoare improbabile, chiar dacă dinamic posisbile. Introducerea legii a doua a termodinamicii ne asigură că evoluția viitoare a sistemului modelat este într-adevăr 'tipică' și elimină variantele atipice care nu au relevanță pentru problema în chestiune. Cu ajutorul legii a doua devine posibil să calculăm evoluțiile viitoare ale sistemului în care sunt prea multe particule pentru ca un tratament al mișcărilor individuale să fie pus în practică.

Este o întrebare interesantă — și nu lipsită de profunzime — de ce aceste evoluții nu pot fi descrise înspre *trecut*, în ciuda faptului că ecuațiile dinamice ale lui Newton, Maxwell și Einstein sunt toate complet simetrice în timp; în lumea reală legea a doua a termodinamicii nu se aplică în direcția contrară a timpului. Motivul esențial pentru acest lucru pare să aibă de-a face cu condițiile foarte speciale care au avut loc la începutul timpului — la originea Universului (vezi Capitolul 7 din MN). De fapt, aceste condiții inițiale au fost atât de speciale încât

ele furnizează un alt exemplu al preciziei extraordinare cu care comportamentul fizic observat este modelat de ipoteze matematice definite.

În cazul *Big-Bang*ului, o parte esenţială a ipotezei relevante este că în fazele sale timpurii, conţinutul material al Universului a fost într-o stare de *echilibru* termic. Ce înseamnă asta? Studiul stărilor de echilibru termic reprezintă extrema opusă în raport cu modelarea precisă a mişcărilor detaliate a câtorva obiecte, ca în cazul discutat al pulsarului.Acum nu ne mai interesează decât comportamentul tipic în sensul cel mai exact al cuvântului. O stare de echilibru, în general, este starea unui sistem care 's-a liniştit' de tot şi de la care sistemul nu va devia semnificativ chiar dacă este sub influenţa unor mici perturbaţii. Pentru un sistem cu un mare număr de particule (un mare număr de grade de libertate) — astfel încât nu mai poate fi vorba despre mişcări ale particulelor individuale şi de comportamentul mediu şi măsuri mediate. cum r fi temperatura sau presiunea — starea de echilibru termic este starea spre care sistemul va converge în cele din urmă în acord cu legea a doua a termodinamicii (sistemul va ajunge la entropie maximă). Calificarea 'termic' implică existenţa unui anumit tip de mediere asupra unui număr mare de mişcări individuale ale particulelor implicate. Termodinamica este cea care se ocupă de aceste medieri — adică cu comportamentul tipic în locul celui individual.

Mai precis vorbind, atunci când ne referim la starea termodinamică a unui sistem sau la echilibru termic, acest lucru nu se aplică unei stări individuale, ci mai degrabă unui ansamblu de stări, care pe o scală macroscopică apar indiscernabile (iar entropia este, în limi mari, logaritmul numărului de stări din acest ansamblu). În cazul unui gaz aflat în echilibru, dacă fixăm presiunea, volumul şi cantitatea şi compoziţia particulelor de gaz, obţinem o distribuţie specifică a vitezelor probabile ale particulelor la echilibru termic (aşa cum a fost prima oară descrisă de Maxwell). O analiză mai detaliată arată o scală pe care sunt aşteptate fluctuaţii statistice care îndepărtează sistemul de starea de echilibru — iar acum intrăm într-un domeniu de studiu al comportamentului statistic al materiei ceva mai sofisticat, care se numeşte *mecanică statistică*.

Din nou, pare că nu există nimic esenţialmente non-computabil în legătură cu modelarea comportamentului fizic cu ajutorul structurilor matematice. Atunci când calculele corespunzătoare au fost făcute, vom găsi o bună corespondenţă între ceea ce este calculat şi ceea ce este observat. Totuşi, atunci când considerăm

sisteme ceva mai complicate decât gaze diluate sau colecţii mari de corpuri gravitând unele spre altele, s-ar putea ca lucrurile să nu mai fie atât de simple, mai ales dacă intră în discuţie natura cuantică a obiectelor implicate. Chiar şi în cel mai pur exemplu de comportare termodinamică — starea de echilibru termic a materiei şi radiaţiei numită stare de corp negru — nu putem folosi exclusiv o tratare clasică, ci suntem obligaţi să vedem că procesele cuantice sunt în mod fundamental implicate. De fapt, tocmai analiza pe care Max Planck a făcut-o radiaţiei corpului negru în 1900 a iniţiat întregul subiect al teoriei cuantice.

Cu toate acestea, predicţiile teoriei fizice (acum ale teoriei cuantice) sunt verificate la modul triumfător. Relaţia observată între frecvenţă şi intensitatea radiaţiei la acea frecvenţă dă o concordanţă foarte bună între experimente şi formula matematică propusă de Planck.

În ciuda faptului că acest subcapitol a fost consacrat naturii computaţionale a teoriei *clasice* nu pot să rezist în acest punct să nu vă arăt ceea ce este de departe cel mai bun exemplu despre care ştiu, de potrivire între observaţie şi formula lui Planck. Acest exemplu ne furnizează de asemenea o minunată confirmare observaţională a modelului standard al Big-Bangului în raport cu ceea ce susţine că ar trebui să fie condiţiile termice precise în care Universul ar fi trebuit să se găsească în primele momente ale existenţei sale. În figura 4.12, punctele individuale indică valorile observate ale intensităţii radiaţiei cosmice de fond la diferite frecvenţe, aşa cum au fost observate de satelitul COBE; curba continuă este desenată în acord cu formula lui Plank, luând temperatura radiaţiei de (pentru cea mai bună potrivire) 2,735 (+/-0,06)K. Precizia cu care aceste rezultate concordă este extraordinară.

Exemplele pe care le-am ales au fost luate din astrofizică, unde comparaţia între .calcule complicate şi comportarea observată a sistemelor naturale este dezvoltată cu precădere. În astrofizică nu putem experimenta direct; de aceea teoriile trebuie testate comparând rezultatele calculelor bazate pe modele (care la rândul lor se bazează pe legile fizice standard) cu observaţii fizice sofisticate (aceste observaţii pot fi făcute de pe pământ, din baloane sau avioane, ca şi din rachete sau sateliţi; ele presupun întotdeauna multe feluri de detectori, pe lângă telescoapele ordinare). Astfel de calcule sunt în mod particular relevante pentru cele ce ne vor interesa în continuare şi le-am menţionat, mai ales, pentru că ele ne furnizează câteva exemple foarte bune asupra modului în care calcule detaliate ne

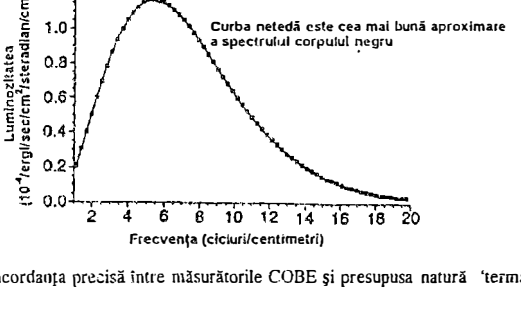


Fig. 4.12. Concordanța precisă între măsurătorile COBE și presupusa natură 'termală' a radiației *big-bang*.

ajută să explorăm Natura, respectiv a modului în care procedurile computaționale pot imita natura. În cele ce urmează ne vor interesa mai curând sistemele biologice. Pentru că, în acord cu cele spuse în partea întâi în comportamentul creierelor conștiente va trebui să căutăm vreun rol pentru acțiuni fizice non-computaționale.

Fără îndoială că modelele computaționale joacă un rol important în modelarea sistemelor biologice, însă aceste sisteme sunt de obicei mult mai complicate decât cele cu care ne întâlnim în astrofizică și deci modelele computaționale vor fi mai puțin credibile. Există foarte puține sisteme destul de 'curate' pentru ca să se ajungă la vreun grad de precizie deosebit. Sisteme comparativ simple, ca de exemplu curgerea sângelui prin diferite tipuri de vase de sânge, pot fi modelate destul de eficient. De asemenea, transmisiunea de-a lungul fibrelor nervoase poate fi modelată — cu toate că în acest caz devine relativ neclar dacă problema rămâne sau nu în cadrul fizicii clasice, deoarece avem de-a face și cu interacțiuni chimice, nu numai cu cele fizice.

Acțiunile chimice sunt un rezultat al efectelor cuantice și astfel părăsim domeniul fizicii clasice ori de câte ori luăm în considerare procese care țin de domeniul chimiei. Cu toate acestea, foarte adesea acțiuni care au la bază efecte cuantice sunt tratate cu ajutorul fizicii clasice. Cu toate că tehnic nu este corect, de cele mai multe ori ideea este că efectele subtile ale teoriei cuantice — dincolo de cele care pot fi subsumate regulilor standard ale chimiei, fizicii clasice și geometriei — nu sunt relevante. Pe de altă parte, mi se pare că deși acest procedeu poate părea rezonabil pentru modelarea multor sisteme biologice (poate chiar și pentru propagarea semnalului nervos), este riscant să tragem din el concluzii

generale asupra acțiunilor biologice mai subtile sau sistemelor biologice mai sofisticate cum este creierul uman. Dacă încercăm să construim teorii despre posibilitatea teoretică a unui model computațional de încredere pentru creierul uman, trebuie mai întâi să explorăm și să înțelegem misterele teoriei cuantice.

În următoarele două capitole voi încerca exact acest lucru – cel puțin în măsura în care este posibil. Unde cred că nu este posibil, voi încerca să argumentez că teoria însăși trebuie amendată pentru a vedea cum s-ar potrivi mai bine cu o imagine credibilă despre lume.

Note și referințe bibliografice:

1. Vezi, de exemplu, Dennett (1991), p.49
2. O astfel de ecuație importantă este prima lege a termodinamicii:

$$dE=tdS-pdV$$

Aici E,T,S,p și V sunt, respectiv, energia, temperatura, entropia, presiunea și volumul unui gaz.

3. De exemplu, Dennett, (1991)
4. Sakharov (1967), cf. Misner etal. (1973), p.428
5. Pentru o descriere grafică, deși nu foarte detaliată a legii a doua a termodinamicii, vezi MN, Cap. 6. Pentru relatări ceva mai sofisticate vezi Davies (1974) și Penrose (1970).

5. STRUCTURA LUMII CUANTICE

5.1. Teoria cuantică — enigmă şi paradox

Teoria cuantică furnizează o descriere superbă a realităţii la scară mică; totuşi, ea conţine multe mistere. Este dificil să-i înţelegem realizările şi, mai mult, este destul de greu să înţelegem tipul de 'realitate fizică' — sau, mai precis, lipsa ei — pe care pare s-o atribuie lumii noastre. Luată aşa cum apare, teoria cuantică pare să ducă la un punct de vedere pe care mulţi — şi printre aceştia mă includ şi pe mine — îl găsesc cât se poate de nesatisfăcător. În cazul cel mai bun, dacă-i luăm descrierile la modul literal, mecanica cuantică ne furnizează o imagine stranie asupra lumii. În cazul cel mai rău, dacă luăm în serios declaraţiile unora dintre cei mai faimoşi protagonişti ai săi, ne lasă fără nici un fel de imagine asupra lumii.

După părerea mea, ar trebui să facem distincţie între două feluri diferite de mistere pe care teoria ni le furnizează. Sunt, mai întâi, cele pe care le voi numi mistere Z, sau mistere *puzzle* (enigme), care ne pot zăpăci, dar care au un suport experimental direct — adevăruri cuantice care privesc lumea în care trăim. În aceeaşi categorie voi include şi acele lucruri care, deşi nu au fost direct experimentate încă, fac parte dintr-un ansamblu de fapte bine stabilite şi avem toate motivele să credem că vor îndeplini prezicerile teoriei cuantice. Unele dintre cele mai izbitoare mistere Z sunt cele denumite fenomene *Einstein-Podolski-Rosen* (sau EPR) şi pe care le voi discuta mai târziu (în 5.4, 6.5). Celălalt tip de mistere cuantice sunt cele pe care le voi denumi mistere X, sau mistere *paradox*. Acestea

sunt lucruri pe care formalismul cuantic pare să le spună despre realitate, dar care au o asemenea natură paradoxală şi neplauzibilă încât ne este greu să credem despre ele că sunt în vreun sens 'actual' adevărate. Ele sunt misterele care ne fac să nu luăm formalismul în serios, să ne îndoim că el ar putea să ne dea o imagine credibilă asupra lumii. Cel mai bine cunoscut mister **X** este paradoxul pisicii lui Schröedinger, în care formalismul mecanicii cuantice pare să ne spună că obiectele mari — pisicile, de exemplu — pot exista simultan în două stări diferite; de exemplu într-o combinaţie nedeterminată de 'pisică vie' şi 'pisică moartă'. (Voi discuta acest tip de paradox în 6.6, 6.9. Vezi figura 6.3 .)

De foarte multe ori, dificultăţile pe care prezentul le are cu teoria cuantică sunt atribuite folosirii unor concepte ale trecutului. Conform acestei teorii, cu cât generaţie după generaţie va deveni familiară cu teoria cuantică şi misterele ei, după un număr suficient de mare de generaţii, oamenii vor fi obişnuiţi să accepte fără dificultate atât misterele **Z** cât şi misterele **X**. Părerea mea, însă, diferă fundamental de această teorie.

Cred că misterele **Z** sunt lucruri pe care ne vom obişnui să le privim şi să le acceptăm ca naturale. Nu cred că va fi la fel cu misterele **X**. După părerea mea, misterele **X** sunt filozofic neacceptabile şi apar doar pentru că teoria cuantică este o teorie incompletă — sau, mai degrabă, pentru că nu este precisă sau potrivită la nivelul la care misterele **X** încep să apară. Cred că, într-o teorie cuantică îmbunătăţită, misterele **X** vor fi pur şi simplu eliminate de pe lista misterelor cuantice. Cât despre misterele **Z**, va trebui să ne obişnuim cu prezenţa lor!

Acestea fiind spuse, se pot pune o serie de întrebări despre cum am putea trage o linie între misterele **Z** şi misterele **X**. Unii fizicieni ne vor spune că nu există mistere în mecanica cuantică care să poată fi notate cu **X**, în sensul că toate lucrurile stranii şi aparent paradoxale pe care formalismul cuantic ni le pune în faţă trebuie să fie de fapt adevărate — să descrie lumea aşa cum arată ea dacă o privim din unghiul potrivit. (Aceşti oameni, dacă sunt cu adevărat logici în demonstraţie şi dacă iau în serios descrierea prin stări cuantice a realităţii fizice, ar trebui să creadă într-o formă a doctrinei 'lumilor multiple', despre care vom vorbi în paragraful 6.2. După aceasta, pisica vie şi pisica moartă a lui Schrödinger vor locui în două Universuri paralele. Dacă priveşti pisica, atunci vor fi de asemenea două copii ale tale în fiecare din aceste două universuri, văzând o pisică vie, respectiv o pisică moartă.) Alţi fizicieni tind către opinia contrară şi vor susţine că am fost

mult prea generos cu formalismul cuantic atunci când am fost de acord că toate enigmele de tip EPR vor fi sprijinite la un moment dat de experimente. Nu vreau să oblig pe nimeni să preia punctul meu de vedere în ceea ce priveşte demarcaţia dintre misterele **Z** şi misterele **X**. Alegerea mea este guvernată de speranţa că vor fi consistente cu punctul de vedere prezentat în cele ce urmează, în 6.12.

Nu voi face în aceste pagini o relatare completă despre teoria cuantică. În acest capitol voi încerca să vă ofer o descriere rapidă şi, în linii mari, completă a trăsăturilor sale importante, concentrându-mă mai ales asupra misterelor **Z**. În capitolul care urmează vă voi împărtăşi motivele care mă îndeamnă să cred că, din cauza misterelor X, teoria cuantică actuală ar trebui să fie incompletă, în ciuda potrivirii minunate pe care o are cu experienţa. Cititorii care vor dori să ştie mai mult despre mecanica cuantică pot citi capitolul 6 din **MN**, sau, de exemplu, Dirac (1947) sau Davies(1984).

Mai târziu, în cele ce urmează — capitolul 6, paragraful 6.12 — voi prezenta o idee recentă privind proiectele de completare ale teoriei cuantice pe care le cred relevante. Trebuie să vă avertizez de asemenea că această idee diferă semnificativ de cea din **MN**, deşi motivaţiile sunt oarecum similare. În sfârşit, în 7.10 (şi în 7.8) voi da câteva motive interesante pentru a crede că astfel de programe de completare a teoriei cuantice ar putea fi non-computaţionale în felul în care avem nevoie. Pe de altă parte, teoria cuantică standard este non-computaţională numai în sensul în care conţine elemente aleatoare ca parte a procedurii de măsurare. Aşa cum am subliniat în partea întâia (paragrafele 3.18, 3.19), elementele aleatoare în sine nu ne dau tipul de non-computabilitate de care avem în cele din urmă nevoie pentru o înţelegere a mentalului.

Să începem însă cu cele mai izbitoare dintre misterele **Z** ale teoriei cuantice. Le voi ilustra în cele ce urmează cu două enigme amuzante.

5.2. Problema Elitzur-Vaidman de testare a bombelor

Imaginaţi-vă o bombă care are un detonator atât de puternic încât cea mai mică atingere o face să explodeze. Chiar şi un singur foton o face să explodeze. Totuşi, în anumite cazuri bomba nu explodează de loc, adică avem de-a face cu un

rebut. Să mai prsupunem că detonatorul este de fapt o oglindă atașată de nasul bombei, astfel încât dacă un foton de lumină vizibilă este reflectat pe această oglindă, reculul pe care el îl transmite este suficient să miște un întrerupător și bomba va exploda — în afară de cazul în care este un rebut (caz în care întrerupătorul cel sensibil este defect). Mai trebuie să presupunem că atâta vreme cât operăm cu unelte clasice, nu există nici o metodă de a verifica, odată ce bomba a fost asamblată, dacă detonatorul funcționează sau nu (singurul moment în care detonatorul se poate strica este cel al asamblării bombei, altfel toate întrerupătoarele sunt bune). Toate acestea sunt prezentate schematic în figura 5.1.

Să ne imaginăm acum că avem o mare cantitate din aceste bombe (fondurile nu sunt o problemă!), însă bănuim că procentajul de rebuturi este destul de mare. Problema este să găsim o bombă care cu siguranță nu este un rebut.

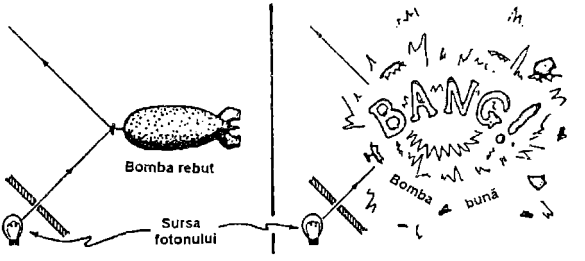


Fig. 5.1. Problema Elitzur Vaidman de testare a bombelor. Detonatorul ultrasensibil al bombei răspunde impulsului unui singur foton - în ipoteza că nu este defect.

Această problemă — și soluția ei au fost propuse de Avshalon Elitzur și Lev Vaidman (1993). Voi amâna explicarea soluției deocamdată. Poate că unii cititori, deja familiarizați cu teoria cuantică și cu ceea ce am denumit mistere **Z**, vor vrea să-și încerce mâna (sau mintea!) și să găsească o soluție. Ajunge să spunem, pentru moment, că **există** o soluție și că această soluție este posibilă pentru tehnologia actuală, cu condiția să avem un număr nelimitat de bombe la dispoziție. Celor care nu sunt încă specialiști în teoria cuantică — sau celor care sunt, dar care nu vor să-și piardă timpul căutând o soluție le cer să aibă răbdare pentru moment (sau să meargă direct la 5.9). Voi ajunge în curând și la soluție, după ce ideile necesare legate de mecanica cuantică vor fi explicate.

În acest stadiu al discuției trebuie să subliniem doar că **există** o soluție;

acest fapt ne dă o măsură a diferenţei între fizica cuantică şi cea clasică. Din punctul de vedere al fizicii clasice nu există nici un mod în care am putea decide dacă detonatorul bombei funcţionează sau nu — altul decât detonarea propriu-zisă, caz în care, dacă bomba nu este un rebut, va exploda şi o vom pierde. Teoria cuantică ne permite ceva cu totul diferit: un efect fizic care rezultă din posibilitatea ca detonatorul să fi fost mişcat — fără ca el să fi fost în mod actual mişcat! Ce este cu adevărat curios în teoria cuantică, este că pot exista efecte fizice actuale care apar din ceea ce filozofii numesc contrafactuale — adică lucruri care s-ar fi putut întâmpla, dar care nu au avut practic loc. În următorul mister **Z**, vom vedea că problema contrafactualelor poate da naştere şi altor tipuri de situaţii.

5.3. Dodecaedrele magice

Pentru a prezenta al doilea mister **Z**, voi începe cu o mică poveste — şi o enigmă[1]. Imaginaţi-vă că am primit de curând un dodecaedru, foarte frumos realizat (figura 5.2). Mi-a fost trimis de o companie cu referinţe superbe, numită 'Tinichelele Chintesenţiale', de pe o planetă a îndepărtatei gigante roşii Betlegeuse. Aceeaşi companie a trimis un dodecaedru similar şi unui coleg de-al meu care trăieşte pe o planetă din sistemul stelei alfa-Centauri, care se află la aproximativ patru ani lumină de noi. Dodecaedrul lui a ajuns acolo cam în acelaşi timp cu al meu. Fiecare dodecaedru are, în fiecare dintre vârfuri, câte un buton. Eu şi cu colegul meu trebuie să apăsăm toate butoanele, pe rând, în mod independent — într-o ordine şi la intervale de timp pe care le putem alege în mod cu totul arbitrar. La o apăsare poate să nu se întâmple nimic (caz în care trebuie să trecem la butonul următor) — sau, la o apăsare putem auzi un clopoţel care sună, acompaniat de un efect pirotehnic de mare amploare care, printre altele, distruge dodecaedrul.

În aceeaşi cutie în care era dodecaedrul am primit o listă a proprietăţilor garantate ale acestuia precum şi instrucţiunile de folosire pentru mine şi pentru colegul meu. Am aflat astfel că trebuie să orientăm dodecaedrele într-un mod foarte precis. Compania oferă indicaţii precise despre cum să aliniem dodecaedrele în relaţie cu, să spunem, centrul galaxiei Andromeda, al galaxiei M-87, etc. Lucrul important este ca dodecaedrul meu şi cel al colegului meu să fie perfect aliniate unul cu altul. Lista instrucţiunilor de folosire este destul de lungă, însă ceea ce

descriu ele este până la urmă destul de simplu.

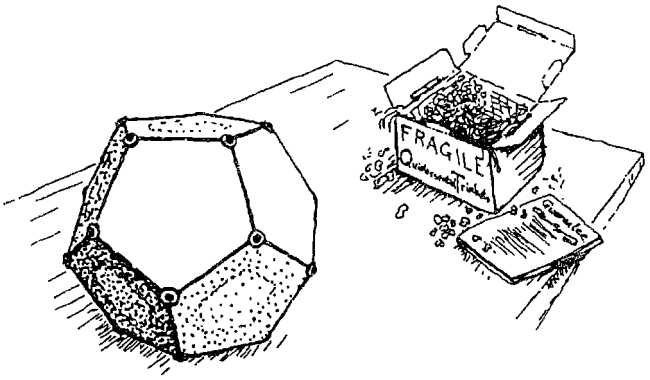


Fig. 5.2 Dodecaedrul magic. Colegul meu are o copie identică pe alfa-Centauri. În fiecare vârf există un buton pe care dacă-l apeşi s-ar putea declanşa clopoţelul şi explozia.

Trebuie să vă reamintesc că 'Tinichelele Chintesenţiale' produc astfel de obiecte de multă vreme să spunem că de peste o sută de milioane de ani — şi nu s-a găsit niciodată vreo greşeală privind proprietăţile pe care ei le garantează. Excelenta lor reputaţie, stabilită în decursul a peste un milion de secole depinde de acest lucru putem fi deci siguri că ceea ce ei pretind este întotdeauna adevărat. Mai mult, există un nemaipomenit premiu în bani (cash) încă nereclamat- pentru oricine găseşte vreo eroare de producţie!

Proprietăţile garantate care ne interesează privesc următoarea secvenţă-tip de apăsare a butoanelor. Colegul meu şi cu mine vom selecta independent unul dintre butoane. Voi numi aceste vârfuri vârfurile SELECTATE. **Nu** apăsăm aceste butoane; ci apăsăm, în schimb, într-o ordine arbitrară la latitudinea noastră, fiecare dintre cele trei butoane de pe vârfurile adiacente celui selectat. Dacă clopoţelul sună la vreuna din aceste apăsări, acest lucru va opri procesul pentru acel dodecaedru particular. S-ar putea însă să nu sune clopoţelul pentru nici una dintre apăsări. Avem nevoie de două proprietăţi (vezi figura 5.3):

a) Dacă eu şi cu colegul meu alegem vârfuri diametral opuse, pe care le vom eticheta SELECTAT, atunci soneria va suna la butonul pe care eu îl apăs

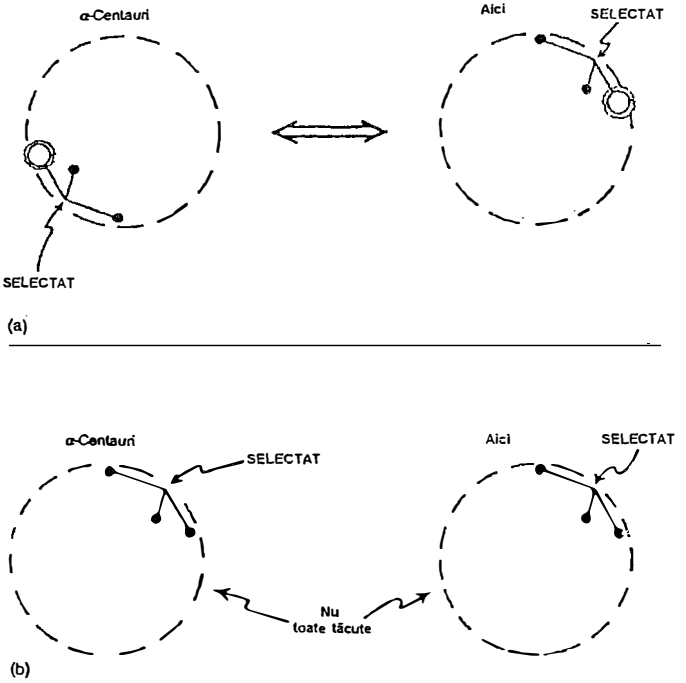


Fig. 5.3. Proprietăţile garantate de firma Tinichelele Chintesenţiale. Dacă SELECTĂM vârfuri opuse, clopoţelul poate suna doar la butoanele diametral opuse...

(adiacent celui SELECTAT) dacă şi numai dacă la colegul meu soneria va suna la apăsarea unui buton diametral opus — indiferent de ordinea particulară în care fiecare dintre noi poate alege să apese butoanele.

b) Dacă am ales amândoi vârfuri corespondente (adică în aceeaşi direcţie pornind de la centru) ca vârfuri SELECTATE, atunci soneria trebuie să sune la cel puţin una din cele şase apăsări pe care le putem efectua în total.

Acum voi încerca să deduc câte ceva despre regulile pe care propriul meu

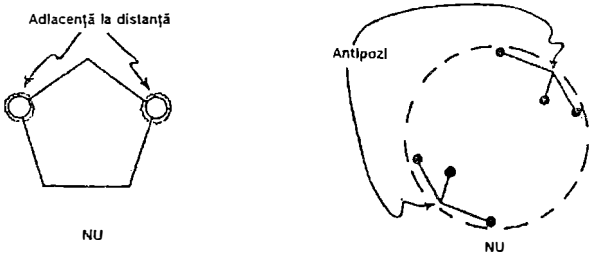


Fig. 5.4. Presupunerea că cele două dodecaedre sunt obiecte total independente (neconectate) ne conduce la concluzia că butoanele unuia sunt predefinite ca *sonore* sau *tăcute*.

dodecaedru trebuie să le satisfacă, independent de ceea ce se întâmplă pe alfa-Centauri şi bazându-mă pe faptul că indicaţiile de folosire a dodecaedrelor magice sunt redactate fără să se ştie ordinea în care butoanele vor fi apăsate. Presupunerea cheie va fi că nu există o 'influenţă' la distanţă care să lege dodecaedrul meu de cel al colegului meu. Astfel, voi presupune că cele două dodecaedre se comportă ca separate, complet independente. Deducţiile mele sunt următoarele (figura 5.4.):

c) Fiecare din vârfurile dodecaedrului meu poate fi desemnat fie ca declanşator de sonerie (Îl vom numi ALB), fie ca tăcut — NEGRU. 'Culoarea' vârfului este independentă de faptul că acest buton este primul, al doilea sau al treilea buton apăsat adiacent celui SELECTAT.

d) Butoanele neadiacente (din două în două) nu pot fi ambele declanşatoare de sonerie (ALBE).

e) Seturile de şase vârfuri adiacente unei perechi de butoane antipodale nu pot fi toate tăcute (adică NEGRE) în acelaşi timp.

(Termenul antipodal se referă la vârfurile diametral opuse pe acelaşi doadecaedru).

Am dedus c) din faptul că i se poate întâmpla colegului meu să aleagă, drept buton SELECTAT, butonul diametral opus celui pe care eu l-am ales; cel puţin, 'Tinichelele Chintesenţiale' nu au nici un mod de a şti că el n-o va face (contrafactualele de care vorbeam înainte!). Astfel, dacă unul din cele trei butoane pe care le aleg va declanşa soneria, atunci trebuie ca vârful diametral opus, dacă va fi apăsat de colegul meu printre primele trei, să declanşeze de asemenea soneria sa; şi aceasta independent de ordinea pe care eu o aleg pentru a apăsa primele trei buroane. Astfel, cum nu există vreo 'influenţă' între cele două dodecaedre, putem

fi siguri că 'Tinichelele Chintesențiale' trebuie să fi aranjat lucrurile în așa fel încât acel vârf particular să declanșeze soneria indiferent de ordinea în care eu apăs butoanele, astfel încât să nu existe un conflict cu a).

De asemenea, d) rezultă din a). Să presupunem că tot al doilea vârf pe dodecaedrul meu declanșează soneria. Oricare dintre acestea îl voi alege prima dată, va declanșa soneria — și să presupunem că am ales vârful adiacent lor ca SELECTAT. Ordinea în care le aleg acum este foarte importantă — ceea ce contrazice a) dacă colegul meu alege din întâmplare butonul lui SELECTAT diametral opus celui ales de mine (o eventualitate pentru care 'Tinichelele Chintesențiale' erau cu siguranță pregătiți).

În sfârșit, e) rezultă din b), împreună cu ceea ce am stabilit până acum. Să presupunem că de data asta colegul meu a ales, ca buton SELECTAT, vârful corespondent celui pe care l-am ales eu ca SELECTAT. Dacă nici unul din cele trei butoane adiacente acestuia nu este un declanșator de clopoțel, atunci, din b) rezultă că unul dintre cele trei vârfuri adiacente pe care colegul meu le apasă trebuie să declanșeze soneria. Rezultă din a) că vârful care pe dodecaedrul meu este opus celui declanșator de sonerie al colegului meu, este la rândul lui declanșator de sonerie — și e) este astfel stabilit.

Acum însă urmează enigma. Încercați să colorați fiecare din vârfurile dodecaedrului fie ALB fie NEGRU în funcție de regulile d) și e). Veți descoperi că, indiferent cât o să vă bateți capul, nu veți reuși. Astfel, o enigmă chiar mai incitantă ar fi să demonstrați de ce nu se pot colora vârfurile după regulile stabilite. Pentru a-i lăsa fiecărui cititor curios șansa să încerce, voi amâna demonstrația mea până la Anexa B, unde voi da o demonstrație destul de directă a faptului că nu putem colora vârfurile dodecaedrului în acest fel. Poate că vreunul dintre cititori va găsi o demonstrație mai ingenioasă!

S-ar putea întâmpla, deci, ca pentru prima dată într-un milion de secole, 'Tinichelele Chintesențiale' să fi făcut o greșeală? Stabilind că este **imposibil** să colorez vârfurile în funcție de regulile c), d) și e) și amintindu-mi de imensul premiu în bani pe care compania îl oferă, vom aștepta cu nerăbdare cei câțiva ani necesari pentru ca mesajul colegului meu, cu descrierea experimentului său, să sosească. Din păcate, la sosirea acestui mesaj, toate speranțele mele privind premiul se spulberă. 'Tinichelele Chintesențiale' au avut dreptate din nou!

Ceea ce arată demonstrația din Anexa B este că nu există, pur și simplu,

nici un mod de a construi — folosind modele clasice — un dodecaedru magic cu proprietăţile garantate de 'Tinichelele Chintesenţiale', dacă considerăm cele două dodecaedre ca obiecte separate independente. Nu este posibil să garantăm cele două proprietăţi cerute, a) şi b), fără un anumit fel de 'conexiune' misterioasă între dodecaedre — o conexiune care persistă până când vom începe să apăsăm pe butoane şi care pare să acţioneze instantaneu la o distanţă de aproximativ patru ani lumină. Şi totuşi, compania 'Tinichelele Chintesenţiale' este în stare să ne ofere garanţia unui lucru care pare imposibil — şi nu au greşit niciodată!

Cum procedează deci 'Tinichelele Chintesenţiale' (sau TC, prescurtarea cunoscută a firmei)? Desigur, TC reprezintă şi prescurtarea pentru teoria cuantică! Ceea ce TC a făcut a fost să aranjeze ca un atom de spin 3/2 să fie suspendat în centrul fiecărui dodecaedru. Aceşti doi atomi au fost produşi pe Betlegeuse într-o stare iniţială combinată de spin total zero, iar apoi au fost separaţi şi izolaţi în centrele celor două dodecaedre astfel încât valoarea totală a spinului să rămână zero (vom vedea ce înseamnă toate acestea în paragraful 5.10). De fiecare dată când eu sau colegul meu apăsăm un buton pe dodecaedrul nostru, se efectuează un anumit tip de măsurătoare (parţială) a spinului, în direcţia dată de distanţa de la centrul dodecaedrului la acel vârf particular. Dacă rezulatul acestei măsurători este găsit afirmativ, atunci soneria va suna, urmată de efectul pirotehnic prevăzut. Voi reveni cu detalii asupra acestui proces de măsurare în paragraful 5.18 şi voi arăta acolo şi în Anexa B, de ce regulile a) şi b) sunt o consecinţă a regulilor standard ale mecanicii cuantice.

Până una alta, concluzia remarcabilă este că presupunerea inexistenţei 'influenţei' la mare distanţă este de fapt violată în mecanica cuantică! O privire aruncată diagramei spaţio-temporale din figura 5.5 vă va lămuri că procesele de apăsare pe buton pe care eu şi colegul meu le efectuăm sunt evenimente separate spaţial (vezi paragraful 4.4). Astfel, conform teoriei relativităţii, nu pot exista semnale care să treacă de la unul la altul şi care să transmită informaţii despre care buton anume este apăsat sau care buton anume declanşează soneria. Totuşi, teoria cuantică susţine că există un anumit fel de 'influenţă' care conectează cele două dodecaedre pe deasupra evenimentelor separate spaţial. De fapt, această 'influenţă' nu se poate folosi pentru a trimite nici un fel de informaţie direct accesibilă într-un mod instantaneu; de aceea nu există un conflict operaţional între teoria cuantică şi teoria relativităţii speciale. Cu toate acestea, există un conflict cu *spiritul*

relativității speciale; ne găsim astfel în prezența unuia dintre acele profunde mistere **Z** ale teoriei cuantice despre care vorbeam: fenomenul *non-localității cuantice*. Cei doi atomi aflați în centrul dodecaedrelor constituie ceea ce se numește o *stare non-separabilă* (*entangled state*) și, după regulile teoriei cuantice, pur și simplu, nu pot fi considerați ca două obiecte independente.

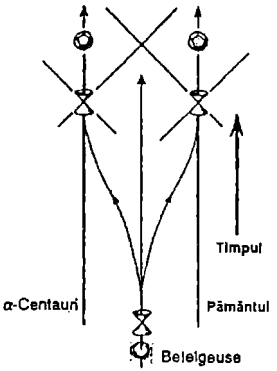


Fig. 5.5. Diagrama spațiu-timp a istoriei celor două dodecaedre. Ele ajung pe Alfa-Centauri și pe Pământ ca evenimete separate spațial.

5.4. Statutul experimental al misterelor Z de tip EPR

Exemplul specific pe care l-am dat mai sus este unul dintr-o clasă mai largă de experimente de gândire numite măsurători EPR, după un celebru articol scris în 1935 Albert Einstein, Boris Podolski și Nathan Rosen (vezi paragraful 5.17 pentru o discuție mai detaliată a efectelor EPR). Versiunea originală a articolului nu se referea la spin, ci la anumite combinații ale poziției și impulsului. Ceva mai târziu, David Bohm a oferit o versiune în care rolul important îi revenea spinului. Experimentul său conținea o pereche de particule de spin ½ (electroni, de exemplu)

emise dintr-un punct într-o stare combinată de spin 0. Concluzia asemănătoare a acestor experimente de gândire a fost că o măsurătoare efectuată într-un loc din spaţiu, asupra unuia dintre membrii perechii cuantice, poate să influenţeze în mod instantaneu celălalt membru într-un mod specific, la orice distanţă s-ar afla cele două particule una de alta. În terminologia teoriei cuantice, cele două particule se află într-o *stare de non-separabilitate* unul cu altul. Fenomenul non-separabilităţii cuantice — un mister-Z veritabil — a fost introdus prima dată ca trăsătură distinctivă a teoriei cuantice de către Erwin Schrödinger (1935b).

Mult mai târziu, într-o remarcabilă teoremă publicată în 1966, John Bell a arătat că există anumite relaţii matematice (inegalităţile lui Bell) între probabilităţile asociate variatelor măsurători ale spinului care pot fi efectuate pe oricare două astfel de particule. Aceste relaţii matematice ar fi consecinţele necesare ale faptului că particulele în chestiune ar trebui să fie entităţi independente, aşa cum spune fizica clasică. Însă în teoria cuantică aceste relaţii vor fi violate într-un mod specific, ceea ce deschide drumul unor experimente posibile pentru a testa dacă relaţiile stabilite de Bell sunt într-adevăr violate de către sistemele fizice actuale aşa cum prevede mecanica cuantică (pe când într-o descriere clasică ele sunt satisfăcute cu necesitate) (vezi **MN**, paginile 304-311 pentru exemple de astfel de sisteme).

Ca o ilustrare a ceea ce această non-separabilitate *nu* înseamnă, John Bell obişnuia să dea ca exemplu *şosetele lui Bertlmann*. Bertlmann era un coleg de-al său care purta invariabil şosete de diferite culori. Acesta era un lucru cunoscut despre Bertlmann (cum l-am întâlnit şi eu pe Bertlmann o dată, pot confirma că propriile mele observaţii sunt consistente cu acest fapt). Astfel, dacă cineva se întâmpla să vadă că una din şosetele lui era verde, putea imediat trage concluzia că cealaltă şosetă *nu* era verde. Cu toate acestea, este nerezonabil să presupunem că o misterioasă influenţă propagându-se instantaneu de la şoseta stângă la cea dreaptă făcea ca lucrurile să stea astfel. Cele două şosete sunt obiecte independente şi nu avem nevoie de 'Tinichelele Chintesenţiale' pentru a ne asigura că proprietatea 'a avea şosete diferite' se va menţine. Efectul poate fi uşor aranjat doar de faptul că Bertlmann hotărăşte de la început că şosetele sale vor fi de culori diferite. Sosetele lui Bertlmann nu violează relaţiile lui Bell şi nu există o influenţă la distanţă mare care să le conecteze. Cu toate acestea, în cazul dodecaedrelor magice, nu putem găsi o explicaţie de genul şosetelor lui Bertlmann care să le

explice proprietăţile. De altfel, aceasta a fost toată poanta discuţiei din paragraful exterior.

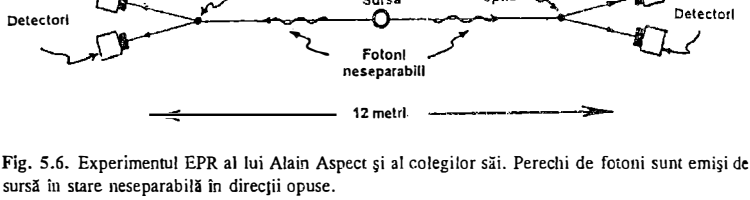


Fig. 5.6. Experimentul EPR al lui Alain Aspect şi al colegilor săi. Perechi de fotoni sunt emişi de sursă în stare neseparabilă în direcţii opuse.

Câţiva ani după ce Bell şi-a publicat articolul, au fost sugerate un număr de experienţe [2], dintre care unele au fost chiar efectuate [3]. Punctul culminant al acestei serii a fost celebrul experiment de la Paris, din 1981, relizat de Alain Aspect şi colegii săi. Ei au folosit perechi de fotoni non-separabili (conform paragrafului 5.17) emişi în direcţii opuse la o distanţă de separaţie de aproximativ 12 metri. Aşteptările teoriei cuantice au fost îndeplinite triumfător, confirmând realitatea fizică a misterelor **Z** de tip EPR, aşa cum au fost ele prezise de teoria cuantică standard şi violând inegalităţile lui Bell (vezi figura 5.6.).

Trebuie menţionat, totuşi, că în ciuda potrivirii foarte precise dintre rezultatele experimentului Aspect şi predicţiile teoriei cuantice, sunt destul de mulţi fizicienii care refuză să accepte că fenomenul non-localităţii cuantice ar fi fost stabilit. Ei spun că detectorii de fotoni din experimentul lui Aspect (şi al celorlalte experimente similare) sunt insuficient de sensibili, astfel că cele mai multe perechi emise de-a lungul timpului de măsurare rămân nedetectate. Argumentul lor este că, dacă detectorii de fotoni ar fi mai sensibili, atunci, într-un anumit stadiu al îmbunătăţirii lor, potrivirea remarcabilă între predicţii şi rezultatele experimentale va dispărea iar relaţiile lui Bell vor fi cumva restaurate.

În ce mă priveşte, cred că este cu desăvârşire improbabil ca excelenta potrivire dintre teoria cuantică şi experiment, care se poate vedea în experimentele Aspect (figura 5.7.), să fie doar un artefact provocat de insensibilitatea detectorilor — şi nu cred că nişte detectori perfecţionaţi vor distruge potrivirea datelor într-un asemenea mod încât relaţiile lui Bell să fie restaurate [4].

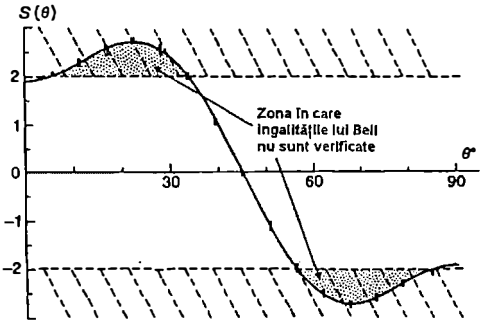


Fig. 5.7. Experimentul Aspect îndeplineşte cu precizie prezicerile teoriei cuantice privind violarea inegalităţilor clasice ale lui Bell. E greu de imaginat că detectori mai buni ar infirma aceste preziceri.

Argumentaţia originală a lui Bell furniza relaţii (inegalităţi) între probabilităţile intersecţiilor diferitelor rezultate posibile. Pentru a estima probabilităţile actuale implicate în experimentul fizic, este necesar să avem un mare număr de observaţii care să fie supuse analizei statistice.

Mai recent au fost propuse un număr de scheme alternative pentru experimente ipotetice de tip da/nu — fără ca în ele să apară probabilităţi. Prima din aceste sugestii recente a fost propusă de Greenberger, Horne şi Zeilinger (1989) şi implică măsurători asupra particulelor de spin ½ în trei locuri separate (să spunem, Pământul, alfa-Centauri şi Sirius, dacă 'Tinichelele Chintesenţiale' ar fi fost implicate în această schemă). Ceva mai devreme, în 1987, Kochen şi Specker au propus o idee oarecum asemănătoare, dar cu particule de spin 1, însă cu configuraţii geometrice foarte complicate; şi Bell însuşi a încercat ceva similar, dar mai puţin explicit, chiar mai devreme, în 1966 (aceste exemple timpurii nu au fost iniţial formulate în termeni de fenomene EPR, însă Heywood şi Redhead, au arătat în 1983 cum se poate face acest lucru; vezi şi Stairs (1983)) [5].

Exemplul meu cu dodecaedre are anumite avantaje deoarece geometria poate fi făcută explicită [6] (există, de fapt, anumite experimente propuse pentru a testa lucruri care sunt echivalente cu aceste diferite exemple de mistere Z, cu toate că într-o formă fizică diferită de cea pe care am prezentat-o aici; vezi Zeilinger *et al.* (1994)).

5.5. Piatra de temelie a teoriei cuantice — o istorie a extraordinarului

Care sunt principiile de bază ale mecanicii cuantice? Înainte de a ne apropia de ele, aş vrea să-mi permit o digresiune istorică. Avantajul va fi evidenţierea celor două ingrediente matematice separate ale teoriei. Este absolut remarcabil şi deloc familiar, că două din elementele fundamentale ale teoriei cuantice moderne au o istorie care începe în secolul şaisprezece, oarecum independent, în opera unui singur om!

Fig. 5.8. Gerolamo Cardano (1501-1976). Medic, inventator, jucător şi matematician. A descoperit probabilităţile şi numerele complexe - cele două elemente de bază ale teoriei cuantice moderne.

Omul în cauză, Gerolamo Cardano (figura 5.8) s-a născut în mizerie (din părinţi necăsătoriţi) la Pavia, în Italia, în 24 septembrie 1501, a fost educat să devină unul din cei mai mari medici ai timpului său şi a sfârşit prin a muri în mizerie la Roma, în 20 septembrie 1576. Cardano a fost un om extraordinar, aproape uitat astăzi. Sper că cititorul mă va ierta dacă voi devia puţin de la subiect

pentru a spune câte ceva despre el, înainte de a mă întoarce la mecanica cuantică propriu-zisă.

Într-adevăr, el nu este cunoscut deloc în mecanica cuantică, cu toate că *numele* său, cel puţin, este bine cunoscut mecanicii automobilelor! Legătura universală dintre cutia de viteze a maşinii şi roţile din spate care permite flexibilitatea cerută pentru a absorbi variatele mişcări verticale ale osiei din spate, se numeşte *arbore Cardan*. Cardano a inventat acest mecanism prin 1545, iar în 1548 a fost în stare să-l încorporeze în trăsura construită pentru împăratul Carol al V-lea, permiţându-i acestuia să călătorească cu mai puţine zdruncinături de-a lungul unor drumuri foarte rele. Cardano a realizat numeroase alte invenţii, ca de exemplu încuietoarea cu cifru, similară celei folosite la casele de bani moderne. Ca medic, a dobândit o mare faimă, având printre pacienţi prinţi şi regi. A fost promotorul multor metode noi în medicină şi a scris o sumedenie de cărţi pe diferite subiecte. Pare să fi fost primul care a observat că bolile venerice cunoscute astăzi sub numele de gonoree şi sifilis sunt două boli separate, cerând tratamente diferite. A propus un tratament de tip 'sanatoriu' pentru bolnavii de tuberculoză cu 300 de ani înainte ca acesta să fie redescoperit de George Boddington, în jurul anului 1830. În 1552, Cardano l-a vindecat pe John Hamilton, Arhiepiscop de Scoţia, de o formă severă de astm — lucru care a afectat însuşi cursul istoriei britanice.

Ce au însă toate acestea de-a face cu teoria cuantică? Absolut nimic, în afară poate de faptul că arată calibrul mental al omului care a descoperit cele două ingrediente separate care vor deveni baza teoriei cuantice. Pentru că în afara calităţilor lui deosebite de medic şi inventator, Cardano a fost un excepţional matematician.

Prima dintre cele două mari descoperiri la care ne referim este *teoria probabilităţilor*. Teoria cuantică este, aşa cum se ştie, mai degrabă o torie probabilistă decât o teorie deterministă. Regulile ei depind în mod fundamental de legile probabilităţii. În 1524 Cardano scrie *Liber de Ludo Aleae* (Cartea Jocurilor de noroc), carte care pune bazele teoriei matematice a probabilităţilor. Cardano a formulat aceste legi cu câţiva ani mai înainte şi le-a aplicat în mod practic la jocurile de noroc, fiind astfel în stare să-şi finanţeze studiile la şcoala medicală din Pavia! Probabil că trebuie să-i fi devenit clar de la o vârstă foarte fragedă că a face bani *trişând* la cărţi este o întreprindere periculoasă, pentru că primul soţ al mamei

sale a fost surprins de un sfârşit neplăcut ca urmare a acestei activităţi. Cardano a descoperit că poate câştiga în mod onest, aplicând descoperirile ce rezultă chiar din legile probabilităţii.

Care este însă celălalt ingredient fundamental al mecanicii cuantice de care vorbeam că a fost descoperit de Cardano? Este vorba de noţiunea de *număr complex*. Un număr complex este un număr de forma $a+ib$, unde '*i*' reprezintă rădăcina pătrată din minus unu ($i = \sqrt{-1}$) şi unde a şi b sunt numere reale obişnuite (numere pe care le scriem acum cu ajutorul unor expresii zecimale). Atunci, a se va numi *partea reală*, iar b , *partea imaginară* a numărului complex $a+ib$. Cardano a ajuns la aceste numere stranii în urma investigaţiilor legate de soluţiile ecuaţiei de gradul trei, deci a ecuaţiei $Ax^3 + Bx^2 + Cx + D = 0$, unde A, B, C şi D sunt numere reale date. În 1545, Cardano a publicat o carte, *Ars Magna*, în care a apărut prima analiză completă a soluţiilor acestor ecuaţii.

Există o poveste nefericită în legătură cu publicarea acestei soluţii. În 1539, un profesor de matematică cunoscut sub numele de Nicolo 'Tartaglia' se afla deja în posesia soluţiei generale a unei anumite clase largi de ecuaţii cubice, iar Cardano a trimis la el un prieten pentru a afla soluţia. Tartaglia l-a refuzat şi Cardano s-a văzut nevoit să se apuce de lucru şi să redescopere soluţia de unul singur. El a publicat rezultatul în 1540, în cartea sa *Practica Aritmeticii şi Măsurătorile simple.* De fapt, Cardano s-a dovedit în stare să extindă soluţia lui Tartaglia pentru a acoperi *toate* cazurile, publicându-şi apoi metoda generală în *Ars Magna.* În ambele cărţi, Cardano a recunoscut întâietatea lui Tartaglia în cazul clasei de ecuaţii pentru care metoda acestuia funcţiona, însă, în *Ars Magna,* a făcut greşeala să declare că Tartaglia i-ar fi dat permisiunea să publice. Tartaglia, furios, pretinde că i-ar fi spus soluţia în cursul unei vizite, cu stricta condiţie să jure că nu o va face niciodată publică. În orice caz, ar fi fost dificil pentru Cardano să-şi publice propriile rezultate fără să menţioneze clasele de soluţii descoperite anterior şi e greu de imaginat cum anume ar fi putut el proceda altfel fără să renunţe pe de-a-ntregul la publicare. Cu toate acestea, Tartaglia i-a purtat o duşmănie ascunsă toată viaţa până când, în 1570, după ce alte evenimente neplăcute zdruncinaseră serios reputaţia lui Cardano, a contribuit din plin la lovitura finală care a marcat sfârşitul unei cariere. Tartaglia a lucrat mână în mână cu inchiziţia, colecţionând un lung dosar cu capete de acuzare împotriva lui Cardano şi contribuind la arestarea şi încarcerarea lui. Cardano a fost eliberat din închisoare numai după ce un emisar

special al Arhiepiscopului de Scoţia (pe care, după cum vă aduceţi aminte, Cardano îl vindecase de astm), s-a deplasat la Roma, în 1571 şi a pledat pentru el, explicând că este vorba despre un 'savant care se îndeletniceşte doar cu buna păstrare şi vindecare a corpurilor în care sufletele date de Dumnezeu să poată trăi până la capătul zilelor lor'

Evenimentele neplăcute la care ne refeream înainte sunt legate de judecarea sub acuzaţie de omucidere a fiului cel mare al lui Cardano, Govanni Battista. La proces, Gerolamo şi-a pus reputaţia în joc pentru a-şi sprijini fiul, lucru care nu i-a fost câtuşi de puţin de folos, pentru că Giovanni era vinovat. El îşi omorâse nevasta — cu care, e adevărat, fusese forţat să se însoare pentru a acoperi un alt omor pe care-l înfăptuise ceva mai demult. Se pare că la omorârea nevestei lui Giovanni a participat şi fiul mai mic, Aldo, încă mai ticălos, care apoi l-a trădat pe Giovanni, ca mai târziu să-şi predea şi propriul tată în mâinile inchiziţiei din Bologna. Răsplata câştigată astfel de Aldo a fost să devină torţionar şi călău public pentru inchiziţia bologneză. Până şi fiica lui Cardano a avut o reputaţie mai mult decât îndoielnică, murind de sifilis, ca rezultat al activităţilor sale profesionale, ca prostituată.

Ar fi un interesant exerciţiu de psihologie istorică să înţelegem de ce Gerolamo Cardano, care pare să fi fost un tată grijuliu, devotat copiilor şi soţiei sale şi care a fost un om onest, sensibil şi cu principii morale, a avut o astfel de descendenţă dezastruoasă. Fără îndoială că atenţia i-a fost deseori distrasă de multiplele lui activităţi şi interese în alte părţi decât la afacerile de familie Din nou fără îndoială, absenţa sa de acasă pentru mai mult de un an după moartea soţiei sale — timp în care îl trata în Scoţia pe arhiepiscop — nu a fost de ajutor dezvoltării copiilor săi. De asemenea, convingerea sa nestrămutată că stelele i-au prevăzut moartea în 1546 — convingere care l-a dus la o activitate febrilă şi neîntreruptă de cercetare şi redactare — trebuie să-l fi determinat să-şi neglijeze soţia, care, în schimb, a murit chiar la sfârşitul acelui an.

Pot să-mi imaginez destul de uşor că soarta nenorocoasă a lui Cardano şi reputaţia sa serios pătată, de eforturile combinate ale copiilor săi, de inchiziţie şi în special de Tartaglia, au fost cauzele datorită cărora este astăzi mult mai puţin cunoscut decât ar merita-o. Pentru mine nu este nici o îndoială că el este una dintre figurile de primă mărime ale Renaşterii. Cu toate că a crescut în mizerie, o atmosferă de studiu a jucat un rol important în anii săi de formare. Tatal său, Fazio

Cardano, a fost geometru; Gerolamo îşi amintea o ocazie în care, la o vârstă fragedă şi-a acompaniat tatăl într-o vizită la Leonardo da Vinci, timp în care cei doi au petrecut lungi ore în noapte discutând chestiuni legate de geometrie.

În ce priveşte episodul publicării rezultatelor lui Tartaglia şi a declaraţiei false că le publică cu permisiunea autorului, trebuie cu siguranţă să respectăm importanţa deciziei de a-şi face publice descoperirile în loc de a le ţine secrete. Pe de o parte, trebuie să nu uităm că mijloacele de trai ale lui Tartaglia depindeau, în mare măsură, de păstrarea secretă a descoperirilor sale (în vederea competiţiilor matematice publice la care lua deseori parte). Pe de altă parte, publicarea acestor rezultate de către Cardano a fost cea care a avut un efect profund şi de lungă durată asupra dezvoltărilor ştiinţei matematice. Mai mult, atunci când în discuţie sunt probleme de prioritate, trebuie să menţionăm că soluţia în litigiu pare să fi fost descoperită (şi) de un alt învăţat al timpului, Scipione del Ferro, profesor la Universitatea din Bologna până în 1526. Sau, cel puţin, del Ferro a fost în posesia soluţiei pe care Tartaglia a descoperit-o mai târziu; nu e foarte clar în ce măsură era conştient de felul în care această soluţie poate fi modificată, pentru a-şi da seama de cazurile considerate mai târziu de către Cardano. Nu există nici o dovadă privind faptul că del Ferro ar fi fost condus către descoperirea numerelor complexe.

Să ne întoarcem însă, cu mai multe detalii, la ecuaţia de gradul trei, pentru a înţelege de ce contribuţia lui Cardano a fost atât de importantă. Nu este greu (cu o substituţie $x \to x+a$) să reducem ecuaţia generală de gradul trei la forma:

$$x^3 = px + q$$

unde p şi q sunt numere reale. Aceasta transformare era bine cunoscută în secolul XVI. Cu toate acestea, trebuie să nu uităm că ceea ce numim acum *numere negative* nu erau acceptate în mod normal ca numere 'propriu-zise', astfel încât diferite versiuni ale ecuaţiei trebuiau scrise în funcţie de diferitele semne ale coeficienţilor p şi q (de exemplu, $x^3 + p'x = q$ sau $x^3 + q' = px$) pentru a păstra toate numerele nenegative care apar în ecuaţie. În cele ce urmează voi adopta notaţia modernă (cea care permite numere negative) pentru a evita complicaţiile excesive.

Soluţiile ecuaţiei de mai sus pot fi exprimate grafic dacă trasăm curbele $y = px + q$ şi $y = x^3$ şi căutăm punctele de intersecţie. Valorile lui x în punctele de intersecţie ne vor da soluţiile ecuaţiei. Priveşte figura 5.9. ecuaţia y

x^3 este reprezentată ca o linie curbă, iar $y = px + q$ este o dreaptă pentru care am indicat diferite posibilități (nu știu dacă Cardano sau Tartaglia au folosit o astfel de descriere grafică, deși ar fi putut-o face; aici ne va fi utilă pentru a vizualiza diferitele situații care pot apărea). Cu această notație, cazurile pe care Tartaglia era în stare să le rezolve apar atunci când p este negativ (sau zero).

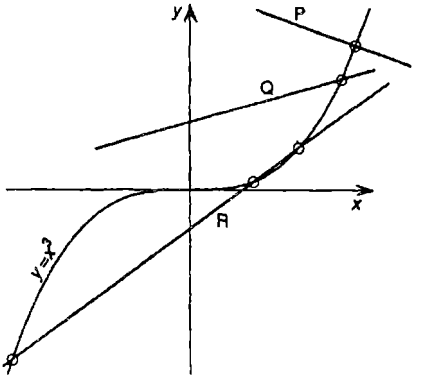


Fig. 5.9. Soluțiile ecuației 'văzute' ca intersecțiile dreptei $y=px+q$ cu $y=x^3$. Cazul Tartaglia corespunde dreptei P, al lui Cardano dreptelor Q, R.

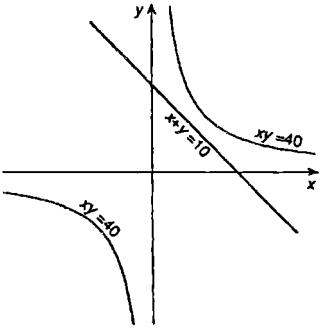


Fig. 5.10. Problema lui Cardano, de a găsi două numere cu produsul 40 și suma 10, revine la a intersecta curba $xy=40$ cu linia $x+y=10$. Ea nu are soluții reale.

În aceste cazuri, linia dreaptă are panta negativă (este îndreptată în jos; un caz tipic este ilustrat de linia P în figura 5.9. Observați că în aceste cazuri există

întotdeauna un singur punct de intersecţie al curbei cu dreapta, astfel încât ecuaţia cubică are exact o soluţie.

În notaţia modernă, putem exprima soluţia lui Tartaglia după cum urmează:

$$x = \sqrt[3]{(w-q/2)} - \sqrt[3]{(w+q/2)}$$

unde

$$w = \sqrt{(q^2/4 + (p')^3/27)},$$

cu $p' = -p$, astfel că expresia conţine doar termeni nenegativi.

Extinderea pe care Cardano o aduce acestei proceduri permite cazurile în care $p > 0$; putem scrie soluţia pentru p pozitivi şi q negativi, însă semnul lui q nu este prea important. Acum, linia dreaptă are panta spre dreapta (însemnată pe figură cu Q sau R).

Observăm că pentru o valoare dată a lui p (adică pentru o pantă dată), dacă $q' = -q$ este destul de mare (astfel încât linia să intersecteze axa Y destul de sus) vom avea din nou exact o soluţie. Expresia pe care Cardano o descoperă pentru aceasta este (în notaţie modernă):

$$x = \sqrt[3]{(w+q'/2)} + \sqrt[3]{(-w+q'/2)}$$

unde

$$w = \sqrt{((q')^2/4 + p^3/27)}$$

Putem vedea, folosind notaţia şi conceptele moderne de numere negative (şi faptul că rădăcina cubică a unui număr negativ este negativul rădăcinii cubice a formei pozitive a acelui număr),' că expresia lui Cardano este principial aceeaşi cu a lui Tartaglia. Totuşi, există ceva cu desăvârşire nou care apare în expresia descoperită de Cardano.

Dacă q' nu este prea mare, dreapta poate intersecta curba în *trei* locuri, astfel încât vom avea trei soluţii ale ecuaţiei originale (două negative, dacă $p > 0$). Acest caz, numit *casus irreducibilis*, apare atunci când

$$(q'/2)^2 < (p/3)^3.$$

Este evident acum că w trebuie să fie *rădăcina pătrată a unui număr negativ*. Astfel, expresiile $q'/2+w$ şi $q'/2-w$ care apar sub semnul radicalului cubic sunt ceea ce vom numi *numere complexe*; însă cele două rădăcini cubice trebuie să dea împreună un număr real pentru a reprezenta soluţii ale ecuaţiei.

Cardano cunoştea foarte bine această problemă misterioasă şi, mai târziu, în *Ars Magna*, adresează explicit întrebarea privind apariţia numerelor complexe în soluţia ecuaţiilor. El ia în considerare problema găsirii a două numere al căror produs este 40 şi a căror sumă este 10, obţinând drept soluţie (corectă) cele două numere complexe

$$5 + \sqrt{-15} \quad \text{şi} \quad 5 - \sqrt{-15}$$

În termeni grafici, asta înseamnă să găsim punctele de intersecţie ale curbei $xy = 40$ cu dreapta $x+y = 10$ (vezi figura 5.10). Să observăm că, aşa cum sunt desenate, liniile din figură nu se intersectează (soluţiile nu sunt numere reale), ceea ce corespunde faptului că avem nevoie de numere complexe pentru a exprima soluţia problemei. Cardano nu era mulţumit cu aceste numere, referindu-se deseori la 'torturile mentale implicate' în lucrul cu ele. Cu toate acestea, necesitatea considerării numerelor de acest fel i-a fost impusă de studiul ecuaţiilor de gradul trei.

Să observăm că avem de-a face cu ceva mult mai subtil, privind felul în care apar numerele complexe în soluţia ecuaţiei cubice (reprezentate în figura 5.9) decât apare problema la prima vedere (adică rezolvea ecuaţiei $x^2-10x+40 = 0$) — în figura 5.10. În cel de-al doilea caz, este clar că nu există o soluţie decât dacă o căutăm printre numerele complexe şi cineva ar putea spune că astfel de numere sunt doar o ficţiune, întrodusă pentru a furniza soluţii unei ecuaţii care nu are de fapt soluţii. Această poziţie, însă, nu este în stare să explice ce anume se întâmplă în ecuaţia cubică. Aici, în *casus irreducibilis* (linia R din figura 5.9.) există trei soluţii *reale* ale ecuaţiei, a căror existenţă nu poate fi negată; şi totuşi, pentru a exprima oricare din aceste soluţii în termeni generali (cu radicali de ordinul doi şi trei în cazul nostru), trebuie să întreprindem întâi o călătorie prin lumea misterioasă a numerelor complexe, până la destinaţia finală, care ne va aduce înapoi în lumea numerelor reale.

Pare că nimeni înaintea lui Cardano nu a văzut cum această lume stă la baza celei pe care o numim reale (alţii, ca de exemplu, Hero din Alexandria şi Diophantus din Alexandria, în primul, respectiv al treilea secol după Christos, par să se fi jucat cu ideea că un număr negativ poate avea un fel de 'rădăcină pătrată', dar nici unul din ei nu a făcut pasul înainte spre combinarea acestor numere cu numerele reale, pentru a ajunge la numerele complexe, nici nu au văzut vreo legătură dintre descoperirile lor şi posibilele soluţii ale unor ecuaţii cu numere

reale).

În cazul lui Cardano, probabil că o curioasă combinație psihologică de misticism și rațională tate științifică, i-au permis să arunce o privire în ceea ce va deveni una dintre cele mai puternice concepții matematice.

În secolele următoare, prin Bombelli, Coates, Euler, Wessel, Argand, Gauss, Cauchy, Weierstrass, Riemann, Levi, Lewy și mulți alții, teoria numerelor complexe s-a dezvoltat într-una din cele mai elegante și mai universal aplicabile structuri matematice.

Dar abia după apariția teoriei cuantice, în primul pătrar al secolului nostru, rolul straniu și atotcuprinzător al numerelor complexe s-a revelat a fi la baza reală a structurii lumii fizice în care trăim și abia astfel a fost percepută legătura profundă dintre numerele complexe și *probabilități*. Nici măcar Cardano n-a putut avea o intuiție a conexiunii misterioase dintre cele două contribuții fundamentale ale sale la matematică — o legătură care formează însăși baza universului material la scara cea mai mică.

5.6. Regulile de bază ale teoriei cuantice

Care este însă această legătură? Cum ajung numerele complexe și teoria probabilităților să se unească și să dea împreună o descriere indiscutabil superbă lucrărilor ascunse ale lumii noastre? Pe scurt, legile numerelor complexe se ocupă de fenomene aflate dincolo de nivelul foarte-micului, iar probabilitățile își joacă partea atunci când este vorba să trecem de la acest nivel la nivelul familiar al percepțiilor noastre — însă va fi nevoie să fiu ceva mai explicit dacât atât dacă e să înțelegem cu adevărat cum stau lucrurile.

Să examinăm mai întâi rolul numerelor complexe. Ele apar într-un fel destul de straniu și greu de acceptat ca aparținând realității fizice. În particular, sunt greu de acceptat deoarece par să nu aibă loc în comportarea fenomenelor perceptibile la nivel normal, acolo unde stăpânesc legile lui Newton, Einstein și Maxwell. Astfel, pentru a ne forma o imagine asupra felului în care lucrează teoria cuantică vom avea nevoie să considerăm deocamdată, în mod provizoriu, că există două nivele distincte ale acțiunii fizice: nivelul fudamental *cuantic*, în care numerele complexe își joacă rolul lor straniu și nivelul *clasic*, al obiectelor

familiare şi al legilor fizice ale lumii macroscopice. Numai la nivel cuantic numerele complexe îşi joacă rolul — un rol care pare să dispară complet la nivel clasic. Asta nu înseamnă că am avea nevoie de o distincţie fizică reală între un nivel la care legile cuantice operează şi un nivel al fenomenelor clasice perceptibile, însă va fi util să ne imaginăm, pentru moment, că există o asemenea diviziune, pentru a fi în stare să înţelegem procedurile care sunt adoptate în teoria cuantică. Dacă există sau nu, în mod *real*, o asemenea diviziune fizică este o chestiune ceva mai profundă care ne va preocupa mai târziu.

La ce nivel începe nivelul cuantic? Trebuie să ne gândim la el ca la un domeniu al lucrurilor 'destul de mici' ca moleculele, atomii, sau particulele fundamentale. Însă această 'micime' nu trebuie să se refere la distanţele fizice. Efecte cuantice pot apărea chiar la distanţe de separare foarte mari. Să ne amintim cei 4 ani lumină care separă cele două dodecaedre din povestirea mea din paragraful 5.3, sau cei 12 metri care separă cele două perechi de fotoni din experimentul Aspect (paragraful 5.4). Nu micimea defineşte deci nivelul cuantic, ci ceva mult mai subtil şi pentru moment este mai bine să nu încercăm să fim expliciţi. Ar fi de folos, totuşi, să ne gândim la nivelul cuantic ca la acel domeniu în care avem de-a face cu foarte mici diferenţe de energie. Mă voi întoarce la acestea mai detaliat în paragraful 6.12.

Nivelul clasic, pe de altă parte, este nivelul experienţei obişnuite, unde funcţionează legile fizicii clasice cu numere reale, iar descrierile obişnuite — ca de exemplu precizarea poziţiei, vitezei şi mărimii unei bile de biliard — au sens. Dacă există sau nu vreo distincţie *actuală* între nivelul cuantic şi nivelul clasic este o întrebare profundă, legată strâns de problema misterelor **X** despre care am amintit în paragraful 5.1. Amânând chestiunea pentru moment, să considerăm deocamdată, printr-o convenţie, că lumea cuantică este separată de lumea clasică.

Care este acum rolul fundamental al numerelor complexe la nivel cuantic? Să ne gândim la o particulă individuală, de exemplu la un electron. În descrierea clasică, electronul poate avea o poziţie A sau o altă poziţie B. Descrierea electronului în mecanica cuantică îi poate lărgi acestuia posibilităţile: el poate avea acum nu numai o locaţie sau alta, dar se poate găsi în oricare dintr-un număr de stări posibile, în care, într-un anumit sens, poate ocupa ambele poziţii simultan! Să folosim notaţia $|A>$ pentru starea în care electronul are poziţia A, respectiv,

notaţia $|B>$ pentru starea în care electronul este în B[1]. În acord cu teoria cuantică, există alte stări posibile pentru electron, notate $w|A> + z|B>$, unde factorii de pondere w şi z care apar aici sunt *numere complexe* (dintre care cel puţin unul trebuie să fie diferit de zero).

Ce înseamnă toate acestea? Dacă factorii de pondere ar fi fost numere *reale* nenegative, am fi putut considera că această combinaţie reprezintă probabilităţile relative ca electronul să fie în A respectiv în B. Atunci, raportul w/z ne-ar fi dat raportul dintre probabilitatea ca electronul să fie în A şi probabilitatea ca electronul să fie în B. Dacă acestea ar fi fost singurele posibilităţi existente pentru electron, speranţa matematică ca electronul să fie în A ar fi fost $w/(w+z)$, iar speranţa ca el să fie în B este $z/(w+z)$. Dacă $w = 0$, atunci electronul ar fi cu siguranţă în B; dacă $z = 0$, va fi cu siguranţă în A. Dacă starea ar fi doar $|A> + |B>$, atunci ar însemna că există probabilităţi egale ca electronul să fie în A sau în B.

Dar w şi z sunt numere complexe, astfel încât această interpretare nu are sens. Rapoartele între ponderile cuantice w şi z nu sunt rapoarte de probabilităţi; şi nu pot fi din moment ce probabilităţile sunt întotdeauna numere *reale*. Nu teoria Cardano a probabilităţilor operează la nivel cuantic, în ciuda opiniei comune după care lumea cuantică este o lume probabilistă. În schimb, teoria misterioasă a *numerelor complexe* este cea care stă la baza unei descrieri matematice precise şi independentă de probabilităţi a nivelului cuantic al lumii.

Nu putem spune, în limbajul obişnuit, ce înseamnă pentru un electron să se afle într-o stare de superpoziţie în două locuri în acelaşi timp, cu factorii de pondere numere complexe w şi z. Trebuie să acceptăm pentru moment acest tip de descriere pentru sistemele nivelului cuantic. Astfel de superpoziţii constituie o parte importantă a construcţiei actuale a micro-lumii şi ne-a fost revelată de către Natură.

Este un *fapt* că am ajuns să descoperim că la nivel cuantic lumea chiar se comportă în mod actual în acest mod straniu şi misterios. Descrierile sunt foarte

---

[1] Folosesc aici notaţia "ket"-Dirac standard pentru stările cuantice; ea se va dovedi convenabilă în cele ce urmează. Cititorii care nu sunt familiari cu formalismul mecanicii cuantice nu trebuie să se îngrijoreze şi nici să-şi pună întrebări asupra semnificaţiei ei.

Paul Dirac a fost unul dintre cei mai extraordinari fizicieni ai începutului de secol. Printre realizările lui a fost o formulare generală a legilor teoriei cuantice şi, de asemenea, generalizarea ei relativistă, implicând "ecuaţia Dirac" pentru electron, pe care el a descoperit-o. Dirac avea o abilitate deosebită de a "adulmeca" adevărul, judecându-şi ecuaţiile în mare măsură în funcţie de calităţile lor *estetice*!

precise şi ele ne furnizează o micro-lume care evoluează după o descriere matematică precisă şi, mai mult, *complet deterministă.*

5.7. Evoluţia unitară (U)

Ce este această descriere deterministă? Este ceea ce voi numi *evoluţie unitară* şi voi folosi litera U pentru a o desemna. Ea este descrisă de ecuaţii matematice precise, însă pentru noi nu este deocamdată important care sunt aceste ecuaţii. Anumite proprietăţi particulare ale lui U sunt tot ce avem nevoie. În ceea ce se numeşte 'descrierea Schrödinger', U este descrisă de aşa numita *ecuaţie a lui Schrödinger, care furnizează* rata de transformare, în raport cu timpul a *stării cuantice* sau *a funcţiei de undă.* Această stare cuantică, deseori notată cu litera grecească ψ (se pronunţă psi) sau $|\psi\rangle$ şi exprimă întreaga sumă ponderată, cu numere complexe ca factori de pondere, a tuturor posibilelor alternative deschise sistemului. Astfel, în exemplul particular la care ne-am referit mai înainte, pentru care alternativele electronului sunt să fie într-un punct A sau într-un alt punct B, starea cuantică $|\psi\rangle$ ar trebui să fie o combinaţie de numere complexe.

$$|\psi\rangle = w|A\rangle + z|B\rangle$$

unde w şi sunt numere complexe (nu amândouă nule). Combinaţia $w|A\rangle + z|B\rangle$ o numim *superpoziţie liniară* a celor două stări $|A\rangle$ şi $|B\rangle$ Cantitatea $|\psi\rangle$ (sau $|A\rangle$ sau $|B\rangle$) se numeşte, de obicei, *vector stare.* Stările cuantice mai generale (sau vectori stare) sunt de forma:

$$|\psi\rangle = u|A\rangle + v|B\rangle + w|C\rangle + \quad + z|F\rangle,$$

unde u, v, z sunt numere complexe (nu toate zero) şi $|A\rangle$, $|B\rangle$, $|F\rangle$ pot reprezenta diferite locaţii posibile pentru o particulă (sau, poate, o altă proprietate a unei particule, ca de exemplu spinul, conform paragrafului 5.10). Mai general, sumele *infinite* sunt permise pentru o funcţie de undă sau un vector de stare (pentru că există o infinitate de posibilităţi privind poziţia unei particule), însă acest tip de problemă nu este important pentru noi aici.

Există totuşi un aspect tehnic al formalismului cuantic pe care *trebuie* să-l menţionez aici. Acesta este că doar *rapoartele* factorilor complecşi de pondere vor avea semnificaţie fizică (voi avea mai multe de spus despre toate acestea mai târziu). Pentru moment, să observăm doar că pentru orice vector de stare $|\psi\rangle$,

fiecare multiplu complex al său, $u|\psi>$, cu u diferit de 0, reprezintă aceeaşi stare fizică ca şi $|\psi>$. Astfel, de exemplu, $w|A> + uz|B>$ reprezintă aceeaşi stare fizică ca şi $w|A> + z|B>$. În consecinţă, doar raportul $w|z$ are semnificaţie, nu şi w sau z luate separat.

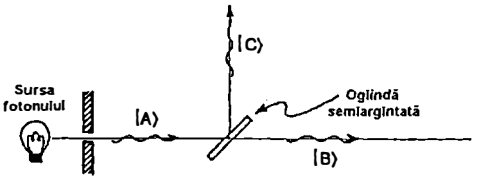


Fig. 5.11. Un foton în starea $|A>$ ajunge pe o oglindă semitransparentă iar starea sa evoluează (prin **U**) într-o superpoziţie $|B> + i|C>$

Trăsătura fundamentală a ecuaţiei Schrödinger (adică a evoluţiei **U**) este *liniaritatea.* Asta revine la a spune că dacă avem două stări, fie $|\psi>$ şi $|\Phi>$ şi dacă ecuaţia Schrödinger ne spune că, după timpul t, cele două stări vor fi, respectiv, $|\psi'>$ şi $|\Phi'>$, atunci orice superpoziţie liniară $w|\psi> + z|\Phi>$ trebuie să evolueze, în acelaşi timp t, spre superpoziţia corespunzătoare $w|\psi'> + z|\Phi'>$. Vom folosi simbolul $\rightarrow$ pentru a nota evoluţia după timpul t. Atunci liniaritatea ne spune că dacă:

$$|\psi> \rightarrow |\psi'> \text{ şi } |\Phi> \rightarrow |\Phi'>$$

atunci evoluţia

$$w|\psi> + z|\Phi> \rightarrow w|\psi'> + z|\Phi'>$$

va fi de asemenea adevărată. Aceeaşi proprietate este de asemenea valabilă pentru superpoziţia liniară a mai mult de două stări cuantice individuale; de exemplu, $u|\chi> + w|\psi> + z|\Phi>$, va evolua în timpul t, în $u|\chi'> + w|\psi'> + z|\Phi'>$, dacă cele trei stări $|\chi>$, $|\psi>$ şi $|\Phi>$ ar evolua independent în $|\chi'>$, $|\psi'>$ şi $|\Phi'>$ Astfel, evoluţia are loc întotdeauna ca şi când fiecare componentă a superpoziţiei ar fi indiferentă la prezenţa altora. Aşa cum le place unora să spună, fiecare 'lume' diferită descrisă de către aceste componente evoluează independent, după aceeaşi ecuaţie Schrödinger deterministă, iar o superpoziţie liniară particulară care descrie întreaga stare, în evoluţie păstrează neschimbate ponderile complexe.

Ţinând cont de toate acestea, ne-am putea imagina că superpoziţiile şi ponderile complexe nu joacă nici un rol fizic efectiv, din moment ce evoluţia

stărilor separate se produce ca şi când alte stări nu ar exista. În realitate, o astfel de interpretare este cum nu se poate mai înşelătoare. Permiteţi-mi să ilustrez printr-un exemplu ce se poate întâmpla.

Să considerăm o situaţie în care lumina se reflectă pe o oglindă pe jumătate argintată — adică o oglindă semitransparentă care reflectă doar jumătate din lumina primită şi transmite cealaltă jumătate. În teoria cuantică lumina este percepută ca o entitate compusă din particule numite *fotoni*. Putem foarte bine să ne imaginăm că pentru un flux de fotoni care cad pe oglindă, jumătate dintre ei sunt reflectaţi, iar jumătate sunt transmişi. De fapt nu este aşa! Teoria cuantică ne spune că, dimpotrivă, *fiecare foton individual*, în momentul în care cade pe oglindă este transpus într-o stare *de superpoziţie liniară* între un proces de reflexie şi unul de transmisie. Dacă, înainte de a întâlni oglinda fotonul este în starea $|A>$, atunci, după ce ajunge la ea, evoluează conform cu U într-o stare care poate fi scrisă $|B> + i|C>$, unde $|B>$ reprezintă starea în care fotonul este transmis prin oglindă, iar $|C>$ starea în care fotonul este reflectat; vezi figura 5.11. Să scriem acest lucru:

$$|A> \rightarrow |B> + i|C>$$

Factorul '*i*' apare pentru că avem o schimbare de fază cu un sfert de lungime de undă [7] care apare între razele reflectată şi transmisă la o asemenea oglindă (pentru a fi mai complet, ar trebui să includ un factor oscilator dependent de timp şi o normalizare peste toţi factorii, însă acestea nu joacă nici un rol în discuţia de faţă în care doresc să intru doar în detaliile esenţiale pentru ceea ce avem deocamdată de demonstrat. Voi spune ceva mai multe despre factorul oscilator în paragraful 5.11 şi voi ridica chestiunea normalizării în 5.12; pentru o descriere mai completă puteţi consulta orice lucrare standard în mecanica cuantică[8]; vezi de asemenea, MN cap.6).

Dacă, pornind de la imaginea clasică asupra unei particule, ar trebui să ne imaginăm că $|B>$ şi $|C>$ sunt doar două lucruri alternative pe care fotonul le poate face, în mecanica cuantică trebuie să încercăm să credem că fotonul face efectiv *ambele lucruri deodată*, în această stranie superpoziţie complexă. Pentru a vedea că nu poate fi vorba doar de o problemă clasică de ponderi de probabilitate, să ducem exemplul mai departe şi să încercăm să aducem cele două părţi ale stării fotonului — cele două raze fotonice — înapoi împreună. Putem face acest lucru provocând mai întâi reflexia fiecărei raze pe o oglindă — complet

argintată de această dată. După relexie [9], starea $|B>$ a fotonului va evolua, în acord cu **U**, într-o altă stare, $i|D>$, pe când $|C>$ va evolua în $i|E>$:

$$|B> \rightarrow i|D> \text{ şi } |C> \rightarrow i|E>$$

Astfel, întreaga stare $|B> + i|C>$ evoluează, prin **U**, în:

$$|B> + i|C> \rightarrow i|D> + i(i|E>) = i|D> - |E>$$

(deoarece $i^2 = -1$). Să presupunem acum că aceste două raze ajung împreună la o a patra oglindă, care este pe jumătate argintată, aşa cum este desenat în figura 5.12 (unde presupun că toate lungimile razelor sunt egale, astfel încât factorul oscilatoriu, pe care îl ignor, continuă să nu joace nici un rol). Starea $|D>$ evoluează într-o combinaţie $|G> + i|F>$, unde $|G>$ reprezintă starea transmisă iar $|F>$ starea reflectată; similar, $|E>$ evoluează în $|F> + i|G>$, deoarece acum starea $|F>$ este cea transmisă iar $|G>$ este cea reflectată:

$$|D> \rightarrow |G> + i|F> \text{ şi } |E> \rightarrow |F> + i|G>$$

Starea totală $i|D> - |E>$ este acum (din cauza liniarităţii lui **U**):

$$i|D> - |E> \rightarrow i(|G> + i|F>) - (|F> + i|G>) =$$
$$= i|G> - |F> - |F> - i|G> = -2|F>$$

(Factorul de multiplicare -2 care apare aici nu joacă nici un rol fizic deoarece, aşa cum am menţionat anterior, dacă starea totală a unui sistem — aici $|F>$ — este multiplicată cu un număr complex diferit de zero, această operaţie lasă situaţia neschimbată). Astfel, vedem că posibilitatea de a fi în starea $|G>$ *nu* este lăsată deschisă pentru foton; cele două raze se combină pentru a produce doar o *singură posibilitate* $|F>$ Acest rezultat curios apare pentru că *ambele raze sunt prezente simultan* în starea fizică a fotonului, între întâlnirile sale cu prima şi cu ultima oglindă. Spunem că cele două raze *interferează* una cu alta. Astfel, cele două lumi alternative ale fotonului, între aceste întâlniri nu sunt cu adevărat separate, ci pot să se afecteze una pe alta prin aceste fenomene de interferenţă.

Este important să ţinem minte că avem de-a face cu proprietatea unui *singur* foton. Fiecare foton individual trebuie să fie considerat ca parcurgând în acelaşi timp ambele drumuri care-i sunt deschise. El rămâne însă *un singur* foton; nu se împarte în doi fotoni în starea intermediară, ci localizarea lui trece printr-un proces neobişnuit de *coexistenţă* a alternativelor descrise de ponderi numere complexe, proces caracteristic teoriei cuantice.

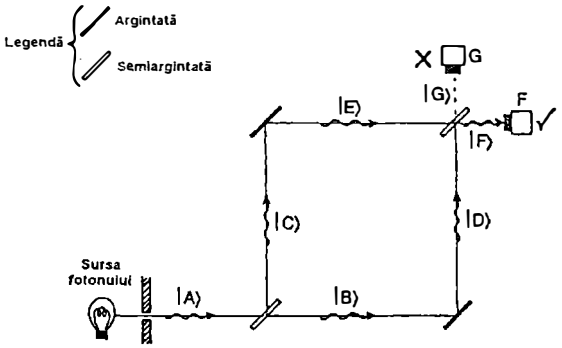


Fig. 5.12. Interferometrul Mach-Zehnder. Cele două părți ale stării fotonului sunt readuse împreună cu două oglinzi complet argintate și o a treia, pe jumătate argintată.

5.8. Reducerea vectorului de stare R

În exemplul considerat mai înainte, fotonul emerge în final într-o stare nesuprapusă. Să ne imaginăm că plasăm detectori (celule fotoelectrice) în punctele marcate cu F și cu G în figura 5.12.

Deoarece, în acest exemplu, fotonul emerge în starea (proporțională cu) $|F\rangle$, fără contribuție de la $|G\rangle$, rezultă că detectorul din F înregistrează fotonul, pe când detectorul din G nu înregistrează nimic.

Ce s-ar întâmpla într-o stare mai generală în care fotonul aflat într-o stare suprapusă ca $w|F\rangle + z|G\rangle$ ajunge să întâlnească ambii detectori? Detectorii noștri fac o *măsurătoare* pentru a vedea dacă fotonul este în starea $|F\rangle$ sau în starea $|G\rangle$. O măsurătoare cuantică are efectul de a 'mări' evenimentele cuantice de la nivelul cuantic la nivelul clasic. La nivelul cuantic, suprapozițiile liniare persistă sub acțiunea continuă a evoluției U. Cu toate acestea, imediat ce efectele sunt 'mărite' la nivel clasic, unde pot fi percepute ca *ocurențe actuale*, nu vom mai găsi lucrurile în combinații ciudate de superpoziții cu ponderi complexe. Ceea ce *găsim*, în acest exemplu, este că fotonul este înregistrat *fie* de detectorul F *fie* de

detectorul G, unde aceste alternative apar cu anumite probabilităţi. Starea cuantică pare să fi 'sărit' în mod misterios dintr-o stare de superpoziţie $w|F\rangle + z|G\rangle$ într-una care este fie $|F\rangle$ fie $|G\rangle$. Această 'săritură' a descrierii stării unui sistem din starea quantică superpusă la o descriere în care au loc una sau alta din alternativele clasice se numeşte *reducţia vectorului de stare* sau *colapsul funcţiei de undă* şi voi folosi litera **R** pentru a desemna această operaţie. Dacă **R** trebuie considerat ca desemnând un proces fizic real sau un fel de iluzie sau aproximaţie este o problemă de care ne vom ocupa în detaliu în cele ce urmează. Faptul că, cel puţin în descrierile noastre matematice trebuie, din când în când, să renunţăm la **U** şi să introducem această procedură complet nouă, **R**, este misterul **X** – fundamental al teoriei cuantice. Pentru moment, e mai bine să nu ne apropiem prea tare de aceste lucruri; deocamdată vom considera **R** ca fiind un proces care se *întâmplă* (cel puţin la nivelul descrierii matematice) ca o trăsătură a procedurii de 'amplificare' a evenimentului de la nivel cuantic la nivel clasic.

Cum calculăm însă probabilităţile diferite pentru rezultatele alternative ale măsurătorii asupra unei stări suprapuse? Există o regulă remarcabilă pentru a determina aceste probabilităţi. Conform acestei reguli, pentru o măsurătoare care decide între stările alternative $|F\rangle$ şi $|G\rangle$, folosind detectorii F şi G, în situaţia de mai sus, atunci, pentru ca detectorii să înregistreze starea

$$w|F\rangle + z|G\rangle$$

raportul dintre probabilitatea ca detectorul F să inregistreze particula şi probabilitatea ca detectorul G s-o facă este dat de:

$$|w|^2 : |z|^2,$$

unde aceste cantităţi sunt modulele pătrate ale w şi z. Pătratul modulului unui număr complex este suma pătratelor părţii sale reale şi imaginare. Deci, pentru

$$z = x+iy$$

unde x şi y sunt numere reale, pătratul modulului este

$$|z|^2 = x^2 + y^2 = (x+iy)(x-iy) = z\bar{z}$$

unde $\bar{z} = x-iy$ este chiar conjugatul complex a lui $z=x+iy$ (în discuţia de mai înainte am presupus implicit că stările pe care le-am notat cu $|F\rangle$, $|G\rangle$ etc. sunt stări normalizate corespunzător. Voi explica aceasta în cele ce urmează; normalizarea este necesară pentru ca această formă a regulii de probabilitate să se păstreze).

Abia aici intră în scenă probabilităţile Cardano. Observăm astfel că

ponderile complexe la nivel cuantic nu joacă rolul probabilităţilor relative (nici n-ar putea, tocmai pentru că sunt numere complexe) dar că pătratul modulelor lor joacă exact acest rol. Mai mult, abia acum — adică la efectuarea *măsurătorii* — avem de-a face cu indeterminare şi probabilităţi. O măsurătoare a stării cuantice are loc atunci când există un proces de amplificare considerabilă a unui proces fizic, o trecere de la nivel cuantic la nivel clasic. În cazul unei fotocelule, înregistrarea unui singur eveniment cuantic, de forma receptării unui foton, poate produce o perturbare la nivel clasic, să spunem de forma unui zgomot, un 'click'. Ca o variantă alternativă am putea folosi o placă fotografică sensibilă pentru a înregistra sosirea unui foton. Aici, evenimentul cuantic al fotonului care atinge placa va fi mărit la nivel clasic sub forma unui semn vizibil pe placă. În ambele cazuri, aparatul de măsură constă dintr-un sistem extrem de sensibil care foloseşte un mic efect cuantic pentru a amorsa un efect mult mai puternic la scară clasică. În această trecere de la nivelul cuantic la nivelul clasic, numerele complexe ale lui Cardano, prin modulele lor pătrate, devin probabilităţile Cardano!

Să vedem cum se aplică această regulă la o situaţie particulară. Să presupunem că, în loc să avem o oglindă în partea din dreapta jos, plasăm acolo o fotocelulă; atunci, această fotocelulă va înregistra starea

$$|B> + i|C>,$$

unde starea $|B>$ provoacă declicul fotocelulei, pe când starea $|C>$ ar trece neînregistrată. Astfel, raportul probabilităţilor respective este $|1|^2 : |i|^2 = 1:1$, ceea ce revine la a spune că probabilitatea ca fiecare dintre evenimente să se producă este aceeaşi; este la fel de probabil ca fotonul să activeze celula ca şi să n-o activeze.

Să considerăm acum un aranjament ceva mai complicat. Să presupunem, în schimb, că în loc să avem fotocelula în partea din dreapta jos, blocăm una dintre razele din exemplul anterior cu un *obstacol* capabil să absoarbă fotonul, să spunem în raza corespunzătoare stării fotonice $|D>$ (figura 5.13); atunci, efectul de interferenţă pe care l-am avut în cazul precedent va fi distrus. *Ar fi* posibilă emergenţa fotonului într-o stare care să implice posibilitatea $|G>$ (pe lângă posibilitatea $|F>$) cu condiţia ca fotonul să *nu* fie absorbit de obstacol. Dacă fotonul *este* absorbit, atunci nu vom avea nici o stare emergentă (combinată sau nu); dacă însă procesul de absorbţie nu are loc, atunci starea fotonului care se apropie de oglinda finală va fi pur şi simplu $-|E>$, care evoluează în $-|F>$

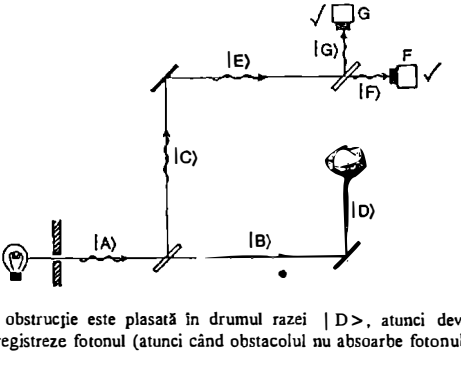


Fig. 5.13. Dacă o obstrucţie este plasată în drumul razei $|D>$, atunci devine posibil pentru detectorul G să înregistreze fotonul (atunci când obstacolul nu absoarbe fotonul!)

$-i|G>$, astfel încât ambele alternative $|F>$ şi $|G>$ sunt într-adevăr incluse în rezultatul final.

În exemplul particular considerat aici, atunci când obstacolul este prezent dar nu absoarbe fotonul, ponderile complexe pentru cele două posibilităţi $|F>$ şi $|G>$ sunt -1 şi $-i$ (starea emergentă fiind $-|F> -i|G>$). Astfel, raportul probabilităţilor respective este $|-1|^2:|-i|^2$, care ne dă din nou probabilităţi egale pentru fiecare din cele două rezultate posibile, astfel încât fotonul va activa cu egală probabilitate detectorul F sau G.

Acum, obstacolul însuşi trebuie de asemenea considerat un 'aparat de măsură': alternativele 'a absorbi fotonul' şi 'a nu absorbi fotonul' fiind două alternative clasice cărora nu li se pot asocia ponderi numere complexe. Chiar dacă obstacolul nu poate fi organizat în aşa fel încât evenimentul cuantic de a absorbi (sau nu) fotonul să poată fi amplificat la un eveniment clasic observabil, trebuie totuşi să considerăm că el ar fi putut fi organizat astfel. Punctul esenţial este că, în urma absorbţiei fotonului, o suprafaţă considerabilă a obstacolului va fi uşor perturbată de către foton, făcând imposibilă colectarea informaţiei conţinută în această perturbare, astfel încât să regăsim efectele de interferenţă ce caracterizează efectul cuantic. Astfel, obstacolul trebuie — cel puţin în practică — să fie considerat ca obiect clasic. El ar juca rolul unui aparat de măsură dacă ar înregistra absorbţia fotonului într-un fel observabil din punct de vedere practic (mă voi întoarce la aceste chestiuni în cele ce urmează, în paragraful 6.6).

Luând în considerare toate acestea, avem libertatea de a folosi 'regula modulelor pătrate' pentru a calcula probabilitatea ca obstrucţia să absoarbă în mod

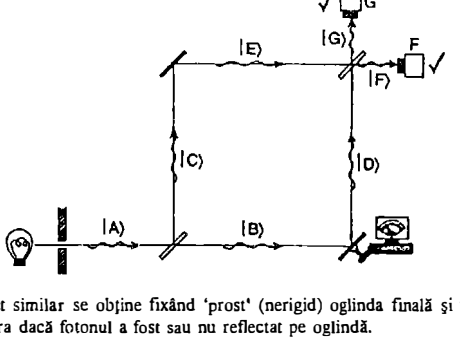


Fig. 5.14. Un efect similar se obţine fixând 'prost' (nerigid) oglinda finală şi folosind 'tremurul' ei pentru a înregistra dacă fotonul a fost sau nu reflectat pe oglindă.

actual fotonul. Starea fotonului atunci când cade pe obstacol este $i|D> - |E>$, iar obstacolul absoarbe fotonul dacă îl găseşte în starea $|D>$ şi nu îl absoarbe în starea $|E>$ Raportul de probabilitate al absorbţiei şi non-absorbţiei este $|i|^2 : |-1|^2 = 1:1$, deci, din nou cele două alternative sunt egal probabile.

Ne-am putea imagina de asemenea o situaţie similară în care, în loc să avem un obstacol în D, putem ataşa un aparat de măsură oglinzii din partea dreaptă jos — în loc să înlocuim oglinda cu un detector, aşa cum am făcut mai devreme. Să ne imaginăm că acest aparat de măsură este atât de sensibil încât ar fi în stare să detecteze (adică să mărească până la nivel clasic) orice perturbaţie transmisă de foton oglinzii atunci când se reflectă pe ea, detecţia fiind semnalată, să spunem, de mişcarea unui ac indicator pe un cadran (vezi figura 5.14). Astfel, starea fotonică $|B>$, la întâlnirea oglinzii, va activa mişcarea acului indicator, însă starea $|C>$ îl va lăsa nemişcat. La întâlnirea cu starea fotonică $|B> + i|C>$, aparatul va 'colapsa funcţia de undă', citind ca stare fie $|B>$ (acul indicator se mişcă), fie $|C>$ (acul rămâne nemişcat), cu probabilitate egală (probabilităţi date de $|1|^2 : |i|^2$).

Procesul **R** va avea deci loc în acest stadiu al experimentului. Pentru a vedea care este comportarea ulterioară a fotonului, putem reface argumentaţia de mai înainte şi vom găsi, aşa cum a fost în cazul obstacolului, că detectorii F şi G vor 'prinde' electronul cu probabilităţi egale (indiferent dacă acul indicator se mişcă sau nu). Pentru acest aranjament experimental, oglinda din partea dreaptă jos trebuie să fie 'prost fixată' (să se poată mişca cu uşurinţă) pentru ca mişcarea acului indicator să poată fi activată; această lipsă de rigiditate a oglinzii va perturba organizarea delicată necesară pentru a asigura 'interferenţa destructivă' între cele

două drumuri ale fotonului între A şi G, cauza pentru care iniţial detectorul din G nu putea să înregistreze.

Cititorul ar putea să observe deja că există un element, nu întru totul satisfăcător referitor la *momentul în care* — şi la *motivul pentru care* — regulile cuantice ar trebui schimbate de la determinismul nivelului cuantic cu ponderi numere complexe, la alternativele non-deterministe ale nivelului clasic cu ponderi probabilistice, caraterizat matematic de considerarea modulelor pătrate ale numerelor complexe implicate.

Ce caracterizează obiectele fizice, precum detectorii F şi G, sau oglinda din dreapta jos — şi de asemenea, materialul absorbant din D — ca fiind *obiecte clasice*, pe când fotonul, la nivel cuantic, să trebuiască tratat atât de diferit? Este doar faptul că fotonul este un sistem fizic simplu, care permite tratarea completă la nivel cuantic, pe când detectorii şi plăcuţele absorbante ar fi sisteme complicate care necesită un tratament alternativ în care subtilităţile mecanicii cuantice sunt eliminate de un tip oarecare de mediere? Mulţi fizicieni vor susţine un astfel de punct de vedere, pretinzând că, riguros vorbind, ar trebui să tratăm *toate* sistemele fizice din punctul de vedere al mecanicii cuantice şi că ar fi o problemă de convenţie să tratăm sistemele mari sau complicate clasic — regulile de probabilitate implicate în procedura '**R**' fiind, cumva, trăsături ale aproximaţiei implicate. Vom vedea în paragrafele 6.6 şi 6.7 că acest punct de vedere nu ne scoate din dificultate — dificultăţile prezentate de misterele **X** ale teoriei cuantice — nici nu explică regula miraculoasă **R** în care probabilităţile apar ca pătratele modulelor ponderilor numere complexe. Pentru moment, totuşi, va trebui să renunţăm la îndoielile noastre; vom continua să investigăm implicaţiile teoriei în special în relaţie cu misterele **Z**.

5.9. Soluţia problemei lui Elitzur şi Vaidman

Suntem acum în situaţia de a putea da o soluţie pentru problema de testare a bombelor din paragraful 5.2. Ideea ar fi să vedem dacă putem folosi oglinda extrem de sensibilă a bombelor ca pe un aparat de măsură, cam în felul în care foloseam obstacolul sau oglinda ne-rigidă, în experimentul anterior. Să considerăm un sistem de oglinzi, două dintre ele pe jumătate argintate, exact în felul în care

am făcut în exemplele de mai înainte, însă cu oglinda bombei jucând rolul de oglindă finală (oglinda din dreapta jos în figura 5.14).

Atunci, ideea este că, dacă bomba e un rebut (în singurul sens pe care îl permitem pentru a nu schimba scopul paradoxului), atunci oglinda este fixată rigid, adică suntem în situaţia descrisă în figura 5.12. Emiţătorul de fotoni trimite un singur foton pe prima oglindă, iniţial în starea $|A>$ Cum situaţia este exact cea din paragraful 5.7, fotonul va emerge în starea (proporţional cu) $|F>$, ca înainte. Deci, detectorul F înregistrează sosirea fotonului, iar detectorul G nu o poate înregistra.

Dacă bomba *nu* este un rebut, atunci oglinda ei este capabilă să răspundă la interacţiunea cu fotonul şi bomba va exploda dacă fotonul ajunge la oglindă. Bomba acţionează deci ca un aparat de măsură. Cele două alternative cuantice 'fotonul ajunge la oglindă' şi 'fotonul nu ajunge la oglindă' sunt mărite până la alternativele clasice 'bomba explodează' şi 'bomba nu explodează'. Cu alte cuvinte, bomba va răspunde stării $|B> + i|C>$ explodând dacă fotonul este în starea $|B>$ şi neexplodând dacă fotonul nu este în starea $|B>$, adică dacă este în starea $|C>$). Probabilităţile relative ale acestor două ocurenţe sunt $|1|^2:|i|^2=1:1$. Dacă bomba explodează, atunci înseamnă că a detectat prezenţa fotonului şi, ceea ce se întâmplă în continuare, nu ne mai interesează. Dacă totuşi bomba nu explodează, atunci starea fotonului este redusă (prin acţiunea lui **R**) la starea $i|C>$, atunci când fotonul ajunge la oglinda din stânga sus şi emerge din acea oglindă în starea $-|E>$ După ce întâlneşte oglinda finală (semitransparentă), starea devine $-|F> -i|G>$ astfel încât există o probabilitate relativă $|-1|^2:|-i|^2=1:1$ pentru cele două rezultate posibile 'detectorul din F întregistrează fotonul' şi 'detectorul din G înregistrează sosirea fotonului' la fel ca în cazurille considerate în secţiunea precedentă când obstacolul nu absoarbe fotonul sau când acul indicator nu se mişcă. Avem astfel posibilitatea ca detectorul G să înregistreze fotonul.

Să presupunem atunci că într-unul din aceste teste găsim că detectorul din G înregistrează cu adevărat fotonul în cazuri în care bomba nu explodează. Din tot ce am spus mai înainte, vedem că acest lucru se poate întâmpla doar dacă bomba *nu* este un rebut! Atunci când avem de-a face cu un rebut, doar detectorul din F poate înregistra. Astfel, în toate cazurile în care G înregistrează, avem o bombă

care cu siguranţă nu este un rebut, ceea ce rezolvă problema de testare[2], din paragraful 5.2.

Se va vedea, din consideraţiile anterioare, că pentru un mare număr de teste, jumătate din bombele active vor exploda şi vor fi deci pierdute. Mai mult, doar în jumătate din cazurile în care o bombă nu va exploda, detectorul G va înregistra fotonul. Astfel, dacă supunem bombele rămase la acelaşi experiment vom avea din nou un număr pentru care G va înregistra şi vom putea repeta procesul din nou şi din nou. În cele din urmă, vom obţine doar o treime ($1/4 + 1/16 + 1/64 + \ldots = 1/3$) din bombele active cu care am început, însă acum sunt toate garantate (nu ştiu la ce vor fi ele folosite în continuare, dar poate e mai prudent nici să nu întreb!).

Vi se poate părea o procedură risipitoare, însă trebuie să vă atrag atenţia că e un lucru remarcabil că poate fi pusă în aplicare chiar şi aşa. Din perspectivă clasică, nu avem nici un mod de a rezolva o astfel de problemă. Doar în teoria cuantică, posibilităţile contrafactuale pot influenţa în mod actual rezultatul fizic. Procedura cuantică pe care am descoperit-o ne permite să realizăm ceva ce părea imposibil — şi *este* într-adevăr imposibil în interiorul fizicii clasice! Mai remarcăm că procedeul poate fi rafinat până la trecerea de la determinarea unei treimi, la determinarea unei jumătăţi din cantitatea de bombe (Elitzur şi Vaidman, 1933). Mai mult, P.G. Kwiat, H. Weinfurter, A. Zeilinger şi M. Kashevich au arătat recent, folosind o procedură diferită, că pierderea poate fi redusă efectiv la zero!

---

[2] Întrerupătorul pentru Sabat.

Faptul că atât Elitzur cât şi Vaidman sunt de la universităţi din Israel ne-a sugerat, într-o conversaţie cu Artur Ekert, un mecanism util celor care păstrează credinţa iudaică tradiţională şi cărora le este interzis să manevreze întrerupătoarele aparatelor electrice în timpul Sabatului. În loc să ne patentăm descoperirea şi să facem avere, am hotărât, în mod dezinteresat, să ne facem ideile publice şi să contribuim la bunăstarea comunităţii evreieşti. Singurul lucru de care avem nevoie este o sursă de fotoni care emite o continuă succesiune de fotoni, două oglinzi pe jumătate argintate, două oglinzi complet argintate şi o fotocelulă plasată în G. Aranjamentul experimental este chiar cel din figura 5.13, cu fotocelula plasată în G. Pentru a activa sau dezactiva aparatul, nu trebuie decât să plasăm un deget în drumul razei în D, în locul obstacolului din figura 5.13. Dacă fotonul ne loveşte degetul, cu aparatul nu se întâmplă nimic şi cu siguranţa nu este vorba de vreun păcat (pentru că fotonii ne lovesc degetele în mod continuu oricum, chiar şi în ziua de Sabat). Dar dacă degetul nu întâlneşte nici un foton, atunci există o şansă de 50% (după Voia lui Dumnezeu) ca întrerupătorul să fie activat. Cu siguranţă nu poate fi un păcat să nu primeşti un foton în deget! (Obiecţia practică care s-ar putea aduce este că sursele care emit fotoni individuali sunt dificil de construit şi scumpe! Dar nu avem cu adevărat nevoie de ele - orice sursă de fotoni va fi bună, deoarece argumentul poate fi aplicat pentru fiecare dintre fotoni în mod individual.)

În privinţa construirii unui aparat experimental care să producă fotoni individuali, trebuie să subliniem că astfel de aparate au fost construite (vezi Grangier *et al.* (1986)).

Iar ca un comentariu final, trebuie să observ că nu este necesar ca aparatul de măsură să fie un obiect atât de dramatic ca bombele de care vorbeam. Un astfel de aparat nici măcar nu trebuie să semnaleze, lumii exterioare, recepţia sau non-recepţia fotonului.

O oglindă nerigidă poate funcţiona ca aparat de măsură dacă este destul de uşoară încât să se mişte semnificativ ca rezultat al impactului fotonului şi să disperseze apoi acestă mişcare ca fricţiune. Chiar faptul că oglinda este nerigidă (cum era cea din experimentul precedent) va permite detectorului G să primească fotonul, chiar dacă oglinda nu se mişcă în mod actual (indicând astfel că fotonul a ales celălţlt drum). Doar *potenţialitatea* mişcării este suficientă pentru a-i permite fotonului să ajungă în G ! Şi obstacolul prezentat în secţiunea anterioară joacă un rol asemănător, servind la 'măsurarea' prezenţei fotonului undeva de-a lungul drumului său descris de stările succesive $|B>$ şi $|D>$ O nereuşită în înregistrarea fotonului, atunci când aparatul respectiv este capabil să-l captureze, contează tot atât de mult ca o 'măsurătoare'

Măsurătorile de acest tip (negativ) se numesc *măsurători nule* (sau libere de interacţiune), vezi Dicke 1981, şi au o importanţă teoretică (şi poate chiar practică) considerabilă. Există experimente pentru a testa direct predicţiile mecanicii cuantice în astfel de situaţii. În particular, Kwiat, Weinfurter şi Zeilinger au reuşit recent un experiment de tipul *precis* care este implicat în rezolvarea problemei de testare Elitzur-Vaidman. Aşa cum de altfel ne-am obişnuit să acceptăm, predicţiile teoriei cuantice au fost complet confirmate.

Măsurătorile nule sunt, într-adevăr, printre cele mai profunde mistere Z ale mecanicii cuantice.

5.10. Teoria cuantică a spinului. Sfera Riemann

Pentru a aborda a doua dintre enigmele mele cuantice introductive, este necesar să privim mai adânc în interiorul structurii teoriei cuantice. Aminteşte-ţi dodecaedrul meu (şi pe cel al colegului meu); ele aveau câte un atom de spin 3/2

în centru. Ce este spinul şi care este importanţa lui în teoria cuantică?

Spinul este o proprietate intrinsecă a particulelor. În principiu este vorba de acelaşi concept fizic ca şi rotaţia sau *momentul de rotaţie* — pentru un obiect clasic cum ar fi o minge de golf, de cricket, sau întregul Pământ. Există, totuşi, o mică diferenţă: pentru obiectele mari, contribuţia majoră la momentul de rotaţie provine de la mişcările orbitale a toturor particulelor una în jurul alteia, pe când pentru o singură particulă, spinul este o proprietate intrinsecă a acelei particule. De fapt, spinul unei particule elementare are trăsătura curioasă că mărimea sa este întotdeauna *aceeaşi*, cu toate că direcţia axei sale de spin poate varia de fapt, această 'axă' se comortă într-un mod foarte ciudat, care nu are nici o legătură cu ceea ce se întâmplă în fizică clasică. Mărimea spinului este descrisă în funcţie de unitatea fundamentală a mecanicii cuantice, $\hbar$, care este simbolul Dirac pentru constanta lui Planck, h, împărţită la 2π. Măsura spinului unei particule este întotdeauna un multiplu întreg (nenegativ) sau semi-întreg de $\hbar$, adică 0, $\frac{1}{2}\hbar$, $\hbar$, $3/2\,\hbar$, $2\hbar$, etc. Spunem că aceste particule au spin 0, spin ½, spin 1, spin 3/2, spin 2, etc.

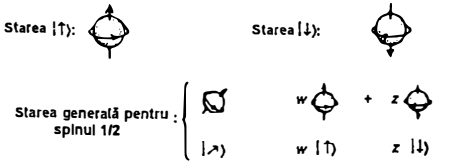


Fig. 5.15. Pentru o particulă de spin 1/2 (electron, proton, neutron), toate stările de spin sunt superpoziţii complexe de două stări: *spin sus* şi *spin jos*.

Să începem prin a considera cel mai simplu caz (lăsând la o parte spinul 0, care este prea simplu: o singură stare de spin, sferic-simetrică), spinul ½, ca de exemplu, cel al electronului sau nucleonului (proton sau neutron). Pentru spinul ½, toate stările de spin sunt superpoziţii liniare de exact *două stări*, să spunem spinul generat de rotaţia spre dreapta: *spin-sus*, vertical, scris $|\uparrow\rangle$, sau spinul generat de aceeaşi rotaţie *jos*, vertical, notat $|\downarrow\rangle$ (vezi figura 5.15). Starea generală de spin ar fi acum o combinaţie de numere complexe

$$|\psi\rangle = w|\uparrow\rangle + z|\downarrow\rangle.$$

De fapt, rezultă că fiecare combinaţie reprezintă starea spinului unei

particule (de magnitudine $\frac{1}{2}\hbar$) într-o anumită direcţie determinată de raportul numerelor complexe w şi z. Nu este nimic special în alegerea acestor două stări $|\uparrow>$ şi $|\downarrow>$ Toate combinaţiile acestor două stări sunt stări de spin la fel de determinate ca cele două originale.

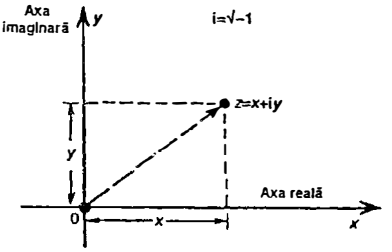


Fig. 5.16. Reprezentarea unui număr complex în planul complex (Wessel-Argand-Gauss).

Să vedem cum putem face această relaţie mai explicită şi geometrică. Acest lucru ne va ajuta să apreciem faptul că ponderile complexe nu sunt nişte lucruri atât de abstracte cum par la prima vedere. De fapt, ele au o relaţie explicită cu geometria spaţiului (îmi închipui că astfel de realizări geometrice l-ar fi încântat pe Cardano şi, probabil, l-ar fi ajutat împotriva 'torturilor mentale' — deşi teoria cuantică ea însăşi ne furnizează destule torturi mentale!).

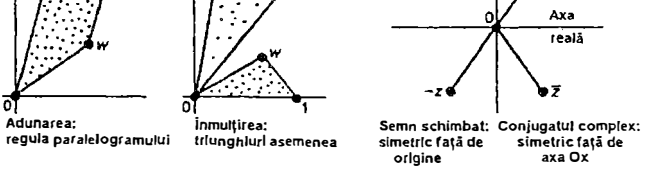


Fig. 5.17. Interpretarea geometrică a operaţiilor cu numere complexe

Va fi util, mai întâi, să considerăm reprezentările standard ale numerelor complexe ca punct în plan (plan numit fie planul Argand, fie planul Gauss, fie planul Wessel, sau doar planul *complex*). Ideea este să reprezentăm numărul

complex $z=x+iy$, unde x şi y sunt numere reale, prin punctul din plan ale cărui coordonate sunt (x,y), raportate la două axe carteziene oarecare (vezi figura 5.16). Astfel, de exemplu, cele patru numere complexe 0, 1, $1+i$, i constituie vârfurile unui pătrat. Există reguli geometrice simple pentru suma şi produsul a două numere complexe (figura 5.17). Negativul $-z$ al unui număr complex z este reprezentat prin operaţia de reflexie în origine; complex conjugatul numărului complex z, notat cu $\bar{z}$, se realizează prin reflexia în raport cu axa reală, (OX).

Modulul unui număr complex este distanţa de la origine a punctului prin care este reprezentat; pătratul modulului este atunci pătratul acestui număr. Cercul unitate este locul geometric al punctelor care se află la distanţa *unu* faţă de origine (figura 5.18); acestea reprezintă numerele complexe ale modulului unitate, numite uneori *fazele pure*, de forma

$$e^{i\theta} = \cos\theta + i\sin\theta ,$$

unde θ este real şi măsoară unghiul pe care îl face cu axa reală (OX) linia care uneşte originea (O) cu punctul care reprezintă numărul complex[3].

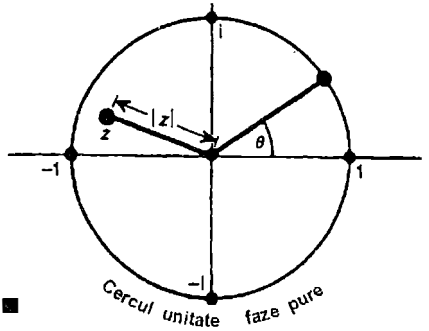


Fig. 5.18. Cercul unitate constituit din numerele complexe $z = e^{i\theta}$, cu θ real, adică $|z|=1$

Să vedem acum cum putem reprezenta *rapoartele* perechilor de numere complexe. În discuţia precedentă am arătat că o stare nu este schimbată din punct

[3] Numărul real e este 'baza logaritmilor naturali': $e=2,7182818285..$; expresia este „e ridicat la puterea z" şi avem:

$$e^z = 1 + z + z^2/(1\times2) + z^3/(1\times2\times3) + z^4/(1\times2\times3\times4) +$$

de vedere fizic dacă este multiplicată cu un număr complex diferit de zero (să ne aducem aminte, de exemplu, că $-2\,|F\rangle$ era aceeaşi stare fizică cu $|F\rangle$ în exemplele noastre anterioare).

Astfel, în general, $|\psi\rangle$ este aceeaşi stare fizică cu $u\,|\psi\rangle$, pentru orice număr complex u diferit de zero. Aplicat stării:

$$|\psi\rangle = w\,|\uparrow\rangle + z\,|\downarrow\rangle,$$

putem vedea că dacă multiplicăm atât w cât şi z cu acelaşi număr complex u diferit de zero, nu vom schimba situaţia fizică reprezentată de acea stare. Raportul celor două numere complexe $z:w$ este cel care este reprezintă diferitele stări de spin fizic distincte ($uz:uw$ fiind la fel ca $z:w$, dacă u este diferit de zero).

Cum reprezentăm geometric un raport de numere complexe? Diferenţa esenţială între un raport complex şi un număr complex oarecare este că *infinitul* este permis ca un raport, pe lângă toate numerele complexe finite. Astfel, dacă ne gândim să reprezentăm raportul $z:w$, în general, printr-un singur număr complex z/w, vom avea probleme când $w=0$. Pentru a acoperi această posibilitate, vom folosi pur şi simplu simbolul ∞ pentru z/w în cazul în care $w=0$. Acest caz apare atunci când considerăm starea particulară 'spin-jos': $|\psi\rangle = w\,|\downarrow\rangle = 0\,|\uparrow\rangle + z\,|\downarrow\rangle$ Să ne amintim că nu este permis să avem atât $w=0$ cât şi $z=0$, însă $w=0$ este permis (putem folosi w/z în loc şi să reprezentăm acest raport dacă vrem; atunci avem nevoie de infinit pentru a acoperi cazul $z=0$, care dă starea particulară 'spin-sus'; nu contează care descriere o folosim).

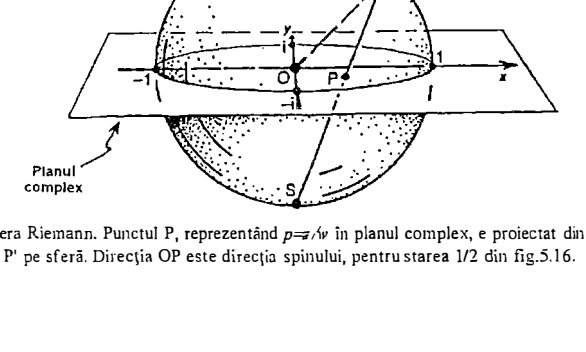


Fig. 5.19. Sfera Riemann. Punctul P, reprezentând $p=z/w$ în planul complex, e proiectat din polul sud S în punctul P' pe sferă. Direcţia OP este direcţia spinului, pentru starea 1/2 din fig.5.16.

Felul în care reprezentăm spaţiul tuturor rapoartelor complexe este să folosim o *sferă*, denumită *sfera Riemann*. Punctele de pe sfera Riemann reprezintă numere complexe sau ∞. Putem reprezenta sfera Riemann ca o sferă de rază unu al cărei plan ecuatorial este planul complex şi a cărei origine este originea planului complex. Ecuatorul sferei va fi identificat cu cercul unitar în planul complex (vezi figura 5.19). Acum, pentru a reprezenta un raport complex particular, să spunem $z:w$, marcăm punctul P în planul complex care reprezintă numărul complex $p=z/w$ (presupunând pentru moment că w nu este zero), iar apoi proiectăm P din plan, într-un punct P' pe sferă, din polul sud S. Asta revine la a spune că luăm linia dreaptă din S în P şi marcăm punctul P' pe sferă ca punct de intersecţie. Această corespondenţă între punctele sferei şi punctele planului se numeşte *proiecţie stereografică*.

Pentru a vedea că este rezonabil pentru polul sud însuşi să reprezinte infinitul, să ne imaginăm un punct P în plan care se mişcă la distanţe foarte mari; atunci, vom găsi că punctul P' care îi corespunde, se apropie de polul sud, atingându-l atunci când P tinde spre infinit.

Sfera Riemann joacă un rol fundamental în imaginea cuantică a sistemelor cu două stări. Acest rol nu e totdeauna explicit, însă sfera Riemann e mereu acolo, în spatele scenei. Ea descrie, într-un mod geometric abstract, spaţiul stărilor fizic distincte care poate fi construit prin superpoziţii liniare, din oricare două stări cuantice distincte. De exemplu, pentru cele două stări care reprezentau locaţiile posibile ale fotonului, să spunem $|B\rangle$ şi $|C\rangle$, combinaţia liniară va avea forma $w|B\rangle + z|C\rangle$

Chiar dacă în paragraful 5.7 am făcut uz în mod explicit doar de cazul particular $|B\rangle + i|C\rangle$, ca un rezultat al reflexiei/transmisiei pe o oglindă pe jumătate argintată, celelalte combinaţii nu trebuie să fie greu de obţinut. Singurul lucru de care avem nevoie este să variem procentul de argintare al oglinzii şi să introducem un segment din mediul refringent în drumul uneia dintre razele emergente. În acest fel, putem construi o sferă Riemann completă a tuturor stărilor alternative, dată de toate situaţiile fizic posibile de forma $w|B\rangle + z|C\rangle$ care pot fi construite din două alternative $|B\rangle$ şi $|C\rangle$

În cazuri ca acestea, rolul geometric al sferei Riemann nu este doar unul aparent. Totuşi, există şi alte tipuri de situaţii, în care rolul geometric al sferei Riemann este manifest. Cel mai clar exemplu apare atunci când stările de spin ale

particulelor sunt ½, ca în cazul electronului şi al protonului. Starea generală poate fi reprezentată ca o combinaţie:

$$|\psi\rangle = w|\uparrow\rangle + z|\downarrow\rangle$$

şi apare (alegând $|\uparrow\rangle$ şi $|\downarrow\rangle$ corespunzător, din clasa de proporţionalitate a posibilităţilor fizic echivalente) că acest $|\psi\rangle$ reprezintă starea de spin, de mărime ½ħ, care este 'dextrogiră' în raport cu axa îndreptată în direcţia punctului de pe sfera Riemann reprezentând raportul z/w. Astfel, orice direcţie din spaţiu joacă rolul unei direcţii de spin posibile pentru orice particulă de spin ½. Chiar dacă cele mai multe stări sunt reprezentate iniţial ca fiind 'combinaţii aiternative misterioase de ponderi numere complexe' (alternativele fiind $|\uparrow\rangle$ şi $|\downarrow\rangle$) putem vedea că aceste combinaţii nu sunt mai misterioase decât cele două stări originale, $|\uparrow\rangle$ şi $|\downarrow\rangle$, cu care am pornit. Fiecare este tot atât de reală ca şi celelalte.

Ce putem spune însă despre stările de spin mai mare? Lucrurile încep să fie puţin mai complicate şi ceva *mai* misterioase!

Descrierea generală pe care o voi da nu este foarte cunoscută fizicienilor de astăzi. Ea a fost imaginată în 1932 de Ettore Majorana, un strălucit fizician italian care a dispărut la vârsta de 31 de ani de pe un vapor care intra în portul Napoli, în circumstanţe încă incomplet elucidate.

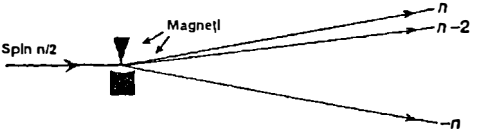


Fig. 5.20. Măsurătoarea Stern-Gerlach. Există $n+1$ rezultate posibile, pentru spin $\frac{1}{2}n$, depinzând cât din spin va fi găsit în direcţia măsurată.

Să considerăm mai întâi ceva *foarte* familiar fizicienilor. Să presupunem că avem un atom (sau 'particulă') de spin $\frac{1}{2}n$. Din nou, vom alege pentru început direcţia 'sus' şi vom pune întrebarea 'cât de mult' din spinul atomului este orientat în mod actual în această direcţie.

Există un aparat standard, numit dispozitivul Stern-Gerlach, care realizează astfel de măsurători folosind un câmp magnetic neomogen. Vom observa că avem exact $n+1$ rezultate posibile, care pot fi distinse prin faptul că atomul este găsit în $n+1$ raze diferite posibile (vezi figura 5.20). Cantitatea de spin care se află pe

direcţia aleasă este determinată de raza particulară în care atomul este găsit. Măsurat în unităţi $\frac{1}{2}\hbar$, cantitatea de spin în această direcţie va avea una din valorile n, $n-2$, $n-4$, $2-n$, $-n$. Astfel, diferitele stări posibile ale spinului, pentru un atom de spin $\frac{1}{2}n$, sunt chiar superpoziţiile (numere complexe) acestor posibilităţi.

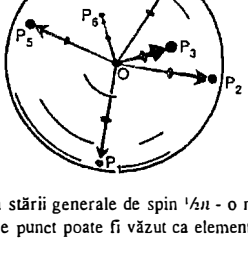


Fig. 5.21. Descrierea Majorana a stării generale de spin $\frac{1}{2}n$ - o mulţime neordonată de n puncte $P_1,\ldots,P_n$ pe sfera Riemann; fiecare punct poate fi văzut ca element de spin $\frac{1}{2}$.

Voi nota diferitele rezultate posibile ale măsurătorilor Stern-Gerlach pentru un spin $n+1$, când direcţia câmpului în aparat este verticală în sus, ca:

$$|\uparrow\uparrow\uparrow \quad \uparrow>,\ |\downarrow\uparrow\uparrow \quad \uparrow>,\ |\downarrow\downarrow\uparrow \quad \uparrow>, \qquad |\downarrow\downarrow\downarrow\ldots\downarrow>$$

corespunzând, respectiv, valorilor de spin n, $n-2$, $n-4$, ..., $2-n$, $-n$, în acea direcţie, unde în fiecare caz sunt exact n săgeţi în total. Ne putem gândi la fiecare săgeată îndreptată în sus ca furnizând o cantitate $\frac{1}{2}\hbar$ de spin în direcţia verticală şi fiecare săgeată în jos furnizând o cantitate $\frac{1}{2}\hbar$ de spin în direcţia verticală în jos. Adunând aceste valori, găsim cantitatea de spin totală, în fiecare caz, obţinută într-o măsurătoare Stern Gerlach orientată în direcţiile sus-jos.

Superpoziţia generală a acestora este dată de combinaţia complexă:

$$z_0|\uparrow\uparrow\uparrow \quad \uparrow> + z_1|\downarrow\uparrow\uparrow \quad \uparrow> + z_2|\downarrow\downarrow\uparrow \quad \uparrow> +\ldots + z_n|\downarrow\downarrow\downarrow\ldots\downarrow>$$

unde numerele complexe z_0, z_1, z_2, z_n nu sunt toate zero. Putem oare să reprezentăm o astfel de stare în funcţie de direcţii ale spinului care nu sunt doar 'sus' sau 'jos'? Ceea ce Majorana a arătat a fost că acest efect e într-adevăr posibil, dar că trebuie să permitem diferitelor săgeţi să fie orientate independent; nu este nevoie ca ele să fie aliniate într-o pereche de direcţii opuse, aşa cum ar fi în cazul măsurătorii Stern-Gerlach. Astfel, vom reprezenta starea generală de spin $\frac{1}{2}n$ ca o colecţie de n astfel de direcţii (ale săgeţilor) independente; ni le putem

imagina ca fiind date de către *n* puncte ale sferei Riemann, unde fiecare direcţie a unei săgeţi este direcţia de la centrul sferei la punctul relevant de pe sferă (figura 5.21). Este important de clarificat că există o colecţie *neordonată* de puncte (sau de direcţii de săgeată). Astfel, nu există o semnificaţie care să fie asociată vreunei ordini a punctelor ca prim, secund, terţ, etc.

Dacă ne imaginăm spinul ca fiind acelaşi tip de fenomen ca cel de la nivelul clasic, vom vedea că descrierea de mai sus ne oferă o imagine destul de ciudată. Spinul unui obiect clasic, ca de exemplu o minge de golf, are o axă bine definită în jurul căruia obiectul se roteşte; în schimb, la nivel cuantic, pare că obiectului îi este permis să se 'rotească' deodată în jurul oricărui tip de axe posibile, în diferite direcţii.

Dacă încercăm să ne gândim la un obiect cuantic ca la unul clasic — exceptând dimensiunile — pare că este obligatoriu să ne confruntăm cu un paradox. Cu cât este mai mare mărimea spinului, cu atât mai multe direcţii sunt implicate. Care este motivul pentru care obiectele clasice nu se rotesc în mai multe direcţii diferite deodată?

Iată un exemplu de mister **X** al mecanicii cuantice. Ceva ajunge să intervină (la un nivel nespecificat) şi găsim că cele mai multe tipuri de stări cuantice nu apar (sau cel puţin, nu apar aproape niciodată) la nivelul clasic al fenomenelor pe care putem să le percepem. În cazul spinului, găsim că doar stările care persistă în mod semnificativ la nivel clasic sunt cele în care direcţiile săgeţilor se aducă în jurul unei direcţii particulare: direcţia (axa) de spin a obiectului clasic care se roteşte.

Există ceva, numit 'principiul de corespondenţă', în teoria cuantică; conform acestuia, atunci când mărimile fizice (ca, de exemplu, mărimea spinului) ajung să fie mari, atunci este posibil pentru sistem să se comporte într-un fel ce aproximează îndeaproape comportamentul clasic (ca în cazul în care săgeţile indică toate aproape aceeaşi direcţie).

Cu toate acestea, principiul de corespondenţă nu ne spune cum pot astfel de stări să apară doar prin acţiunea ecuaţiei Schrödinger **U**. De fapt, 'stările clasice' nu apar aproape niciodată în acest fel. Stările de tip clasic apar din cauza unei proceduri diferite: reducţia vectorului de undă **R**.

5.11. Poziţia şi impulsul unei particule

Există un exemplu şi mai clar în conceptul cuantic de *locaţie* a unei particule. Am văzut că starea unei particule poate implica superpoziţiile a două sau mai multe locaţii diferite (amintiţi-vă discuţia din paragraful 5.7, în care starea unui foton este astfel încât poate fi localizată în două raze simultan, după interacţiunea cu o oglindă pe jumătate argintată).

O astfel de superpoziţie se poate aplica oricărui alt tip de particule — simple sau compuse — ca un electron, un proton, un atom sau o moleculă. Mai mult, nu este nimic în formalismul **U** al teoriei cuantice care să spună că obiecte mai mari, ca de exemplu bilele de biliard n-ar putea să se găsească de asemenea într-o stare suprapusă de mai multe poziţii în acelaşi timp, sau că o minge de golf nu s-ar putea găsi în mai multe stări (suprapuse) de spin în acelaşi timp. Care este explicaţia pentru faptul că anumite obiecte par a fi prea mari, prea masive, sau prea diferite pentru a fi capabile să se comporte ca obiecte 'cuantice' şi nu se găsesc în stări suprapuse în lumea noastră? În teoria cuantică standard, doar acţiunea lui **R** îndeplineşte tranziţia de la nivelul cuantic al superpoziţiei alternativelor posibile la un singur rezultat clasic. Acţiunea lui **U** duce invariabil la superpoziţii clasic 'nerezonabile' (mă voi întoarce la această chestiune în paragraful 6.1).

La nivel cuantic, pe de altă parte, acele stări ale unei particule pentru care nu există o localizare definită pot juca un rol fundamental. Pentru că, dacă particula are un impuls bine definit (adică se mişcă într-un fel definit într-o direcţie anumită şi nu într-o superpoziţie de direcţii diferite în acelaşi timp) atunci starea sa trebuie să cuprindă o superpoziţie a tuturor poziţiilor diferite în acelaşi timp (aceasta este o trăsătură particulară a ecuaţiei Schrödinger; explicarea ei cere însă prea multe detalii tehnice, vezi Dirac (1947), Davies (1984) şi **MN**; ea este asociată de asemenea cu principiul de nedeterminare al lui Heisenberg care ne furnizează limitele în care putem defini poziţia şi impulsul unei particule simultan).

De fapt, stările de impuls bine-definit au un comportament spaţial oscilatoriu, în direcţia de mişcare, care a fost exact trăsătura ignorată în discutarea stărilor fotonice din paragraful 5.7. Vorbind strict, 'oscilator' nu este un termen potrivit, pentru că 'oscilaţiile' nu sunt asemeni vibraţiilor liniare ale unei coarde, în care ne-am putea imagina că factorii complecşi de pondere oscilează înainte şi înapoi prin originea planului complex; ci, dimpotrivă, aceşti factori sunt faze pure

(figura 5.18) care se rotesc împrejurul originii cu o rată constantă — rată care ne dă o frecvenţă proporţională cu energia particulei E, în acord cu faimoasa formulă a lui Planck (vezi figura 6.11 din **MN** pentru o reprezentare 'tirbuşon' a stărilor impulsului). Toate aceastea, chiar dacă sunt importante pentru teoria cuantică, nu vor juca nici un rol particular în discuţiile din continuare; deci cititorul le poate ignora fără nici o problemă.

Mai general, factorii de pondere complecşi nu trebuie să aibă această formă 'oscilatorie', ci pot varia dintr-un punct în altul într-o manieră arbitrară. Factorii de pondere furnizează o funcţie complexă de poziţie numită *funcţia de undă a particulei.*

5.12. Spaţiul Hilbert

Pentru a fi ceva mai explicit (şi mai riguros) privind felul în care acţionează procedura **R** în descrierile standard ale mecanicii cuantice va fi necesar să apelez (dar nu prea mult) la un anumit aparat matematic. Familia tuturor stărilor posibile ale unui sistem cuantic constituie ceea ce este cunoscut sub denumirea de *spaţiu Hilbert.* Nu este necesar să explic ce înseamnă această noţiune intrând în toate detaliile matematice, însă o oarecare înţelegere a spaţiului Hilbert ne va ajuta să clarificăm imaginea pe care o avem despre lumea cuantică.

Prima şi cea mai importantă proprietate care ne interesează este că spaţiul Hilbert e un *spaţiu vectorial complex,* ceea ce înseamnă că ne permite efectuarea combinaţiilor de ponderi numere complexe necesare pentru descrierea stărilor cuantice. Voi continua să folosesc notaţia 'ket' a lui Dirac pentru elementele spaţiului Hilbert. Astfel, dacă $|\psi\rangle$ şi $|\Phi\rangle$ sunt ambele elemente ale spaţiului Hilbert, atunci $w|\psi\rangle+z|\Phi\rangle$ este de asemenea un element, pentru orice pereche de numere complexe w şi z. Aici, este posibil chiar ca $w=z=0$, ceea ce introduce elementul 0 în spaţiul Hilbert pentru acel element al spaţiului care *nu* reprezintă o stare fizică posibilă. Regulile algebrice obişnuite pentru un spaţiu vectorial sunt valabile şi pentru spaţiul Hilbert:

$$|\psi\rangle + |\Phi\rangle = |\Phi\rangle + |\psi\rangle$$

$$|\psi\rangle + (|\Phi\rangle + |\chi\rangle) = (|\psi\rangle + |\Phi\rangle) + |\chi\rangle$$

$$w(z|\psi\rangle) = (wz)|\psi\rangle$$

$$(w+z)|\psi\rangle = w|\psi\rangle + z|\psi\rangle$$
$$z(|\psi\rangle + |\Phi\rangle) = z|\psi\rangle + z|\Phi\rangle$$
$$0|\psi\rangle = \mathbf{0}$$
$$z\mathbf{0} = \mathbf{0}$$

ceea ce înseamnă, mai mult sau mai puţin, că putem folosi notaţia algebrică în felul în care ne-am fi aşteptat.

Spaţiul Hilbert poate avea uneori un număr *finit* de dimensiuni ca în cazul stărilor de spin ale unei particule. Pentru spinul ½, spaţiul Hilbert este doar bidimensional, elementele sale fiind combinaţii liniare complexe ale celor două stări $|\uparrow\rangle$ şi $|\downarrow\rangle$

Pentru spinul ½*n*, spaţiul Hilbert este $(n+1)$ dimensional. Totuşi, uneori, spaţiul Hilbert poate avea un număr *infinit* de dimensiuni — este de exemplu cazul stărilor poziţiei unei particule. Aici fiecare poziţie alternativă pe care particula o poate avea contează ca furnizând o dimensiune suplimentară spaţiului Hilbert. Starea generală care descrie locaţia unei particule cuantice este o superpoziţie complexă a tuturor acestor poziţii individuale diferite (funcţia de undă a paticulei). De fapt, există anumite complicaţii matematice care apar pentru cazul spaţiului Hilbert infinit dimensional; ele însă ar complica inutil discuţia, aşa că mă voi concentra în continuare mai ales asupra cazului finit-dimensional.

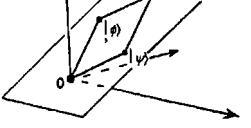


Fig. 5.22. Considerând spaţiul Hilbert ca spaţiu euclidian tridimensional, putem reprezenta suma a doi vectori $|\psi\rangle$ şi $|\Phi\rangle$ cu ajutorul legii paralelogramului (în planul O, $|\psi\rangle$ şi $|\Phi\rangle$).

Atunci când încercăm să vizualizăm un spaţiu Hilbert, întâmpinăm două dificultăţi. În primul rând, aceste spaţii tind să aibă prea multe dimensiuni pentru imaginaţia noastră; în al doilea rând, ele sunt spaţii *complexe* şi nu spaţii reale. Cu toate acestea, este uneori util să ignorăm aceste probleme deocamdată, pentru a dezvolta o intuiţie asupra matematicii. Să spunem, deocamdată, că putem folosi

imaginile bidimensionale şi tridimensionale uzuale pentru a reprezenta un spaţiu Hilbert. În figura 5.22 operaţia de superpoziţie liniară este ilustrată geometric în cazul real tridimensional.

Amintiţi-vă că un vector de stare $|\psi>$ reprezintă aceeaşi stare fizică ca oricare dintre multiplii săi complecşi, $u|\psi>$, unde u este diferit de 0. În termenii imaginii noastre geometrice, asta înseamnă că o situaţie fizică particulară este reprezentată nu printr-un punct în spaţiul Hilbert, ci printr-o întreagă linie — pe care o vom numi rază — care uneşte punctul $|\psi>$ din spaţiul Hilbert cu originea 0. Vezi figura 5.23; trebuie să reţinem însă că, datorită faptului că spaţiile Hilbert sunt complexe şi nu reale, cu toate că o rază arată ca o linie unu-dimensională, este în realitate un întreg plan complex.

Până acum ne-a interesat spaţiul Hilbert doar pentru structura sa de spaţiu vectorial complex. Există însă şi o altă proprietate a sa, aproape la fel de importantă şi care este esenţială pentru descrierea procedurii de reducere **R**. Această proprietate a spaţiului Hilbert este *produsul scalar hermitian* (sau produsul *intern*) care poate fi aplicat oricărei perechi de vectori ai spaţiului Hilbert pentru a produce un număr complex. Produsul intern este o operaţie care ne permite să exprimăm două lucruri importante. Primul este noţiunea de *pătratul lungimii* unui vector în spaţiul Hilbert, adică produsul scalar al vectorului cu el însuşi. O stare *normalizată* (care, aşa cum aminteam în discuţia din paragraful 5.8, ne era necesară pentru ca regula modulului pătrat să fie aplicabilă riguros) este dată de un vector al spaţiului Hilbert a cărui lungime pătrată este egală cu unitatea. Al doilea lucru important pe care ni-l dă produsul scalar este noţiunea de *ortogonalitate* între doi vectori ai spaţiului Hilbert — ei sunt ortogonali dacă produsul lor scalar e zero. Ne putem imagina doi vectori ortogonali ca făcând între ei 'unghi drept'. Importanţa acestui concept în fizica cuantică este considerabilă; diferitele rezulate alternative ale oricărei măsurători sunt ortogonale unul cu altul.

Exemple de stări ortogonale sunt stările $|\uparrow>$ şi $|\downarrow>$ pe care le întâlnim în cazul unei particule de spin ½ (este de observat că ortogonalitatea în spaţiul Hilbert nu corespunde în mod normal noţiunii de perpendicularitate în spaţiu; în cazul spinului ½ stările ortogonale $|\uparrow>$ şi $|\downarrow>$ reprezintă configuraţii fizice care sunt mai degrabă opuse decât perpendiculare). Alte exemple ne sunt furnizate de stările $|\uparrow\uparrow\uparrow...\uparrow>$, $|\downarrow\uparrow\uparrow...\uparrow>$, $|\downarrow\downarrow\uparrow...\uparrow>$, $|\downarrow\downarrow\downarrow...\downarrow>$, pe care le întâlnim în cazul spinului $½n$, fiecare din ele fiind ortogonală cu toate celelalte.

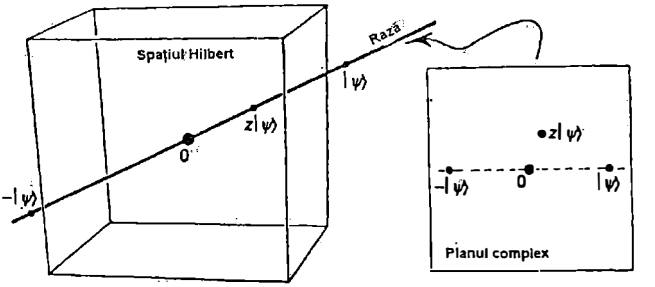


Fig. 5.23. O *rază* în spaţiul Hilbert se compune din toţi multiplii complecşi ai unei stări vectoare $|\psi\rangle$. O privim ca o dreaptă prin originea spaţiului O, dar este de fapt un plan complex.

Sunt de asemenea ortogonale toate poziţiile diferite posibile în care o particulă cuantică se poate afla. Mai mult, stările $|B\rangle$ şi $i|C\rangle$ din paragraful 5.7, care apar la perechile de stări fotonice transmise şi reflectate care apar după ce fotonul a întâlnit oglinda semi-argintată, sunt de asemenea ortogonale, cum sunt şi cele două stări $i|D\rangle$ şi $-|E\rangle$ în care ele evoluează după reflexia pe două oglinzi complet argintate.

Acest ultim fapt ilustrează o proprietate importantă a evoluţiei U , dirijată de ecuaţia lui Schrödinger. Oricare două stări iniţial ortogonale rămân ortogonale dacă fiecare evoluează, în acord cu U, de-a lungul aceleiaşi perioade de timp. Astfel, proprietatea de ortogonalitate este păstrată de către U. Mai mult, U păstrează valoarea produsului scalar dintre stări. Tehnic, de fapt, asta înseamnă termenul de evoluţie *unitară*.

Aşa cum am menţionat anterior, rolul cheie al ortogonalităţii este dat de faptul că, ori de câte ori se efectuează o măsurătoare într-un sistem cuantic, stările cuantice posibile care conduc – după amplificarea la nivel clasic la rezultate care pot fi distinse sunt în mod necesar ortogonale. Acest lucru se aplică, în particular, măsurătorilor nule, ca de exemplu, în problema de testare a bombelor din paragrafele 5.2 şi 5.9. Nedetectarea unei stări cuantice particulare de către un aparat oarecare, capabil să o detecteze, ca fiind rezultatul 'saltului' cuantic într-o stare ortogonală cu cea pe care detectorul este fixat s-o înregistreze.

Aşa cum am mai spus, ortogonalitatea este exprimată matematic ca anularea produsului scalar dintre stări. Acest produs scalar este în general un număr complex atribuit oricărei perechi de elemente ale spaţiului Hilbert. Dacă cele două

elemente (stări) sunt $|\psi\rangle$ şi $|\Phi\rangle$, atunci produsul lor scalar este scris $\langle\psi|\Phi\rangle$. Produsul scalar satisface un număr de proprietăţi algebrice simple pe care le putem scrie astfel:

$$\overline{\langle\psi|\Phi\rangle} = \langle\Phi|\psi\rangle$$
$$\langle\psi|(|\Phi\rangle+|\chi\rangle) = \langle\psi|\Phi\rangle+\langle\psi|\chi\rangle$$
$$(z\langle\psi|)|\Phi\rangle = z\langle\psi|\Phi\rangle$$
$$\langle\psi|\psi\rangle > 0 \text{ dacă } |\psi\rangle \neq \mathbf{0}$$

mai mult, se poate deduce că $\langle\psi|\psi\rangle = 0$ dacă $|\psi\rangle = \mathbf{0}$ Nu vreau să plictisesc cititorul cu prea multe detalii aici (pentru cei interesaţi, recomand oricare carte standard de mecanică cuantică; de exemplu, Dirac (1947)).

Proprietăţile esenţiale de care avem nevoie pentru produsul scalar sunt două:

$|\psi\rangle$ şi $|\Phi\rangle$ sunt *ortogonale* dacă şi numai dacă $\langle\psi|\Phi\rangle = 0$

$\langle\psi|\psi\rangle$ este *pătratul lungimii* lui $|\psi\rangle$.

Să observăm că relaţia de ortogonalitate între $|\psi\rangle$ şi $|\Phi\rangle$ este simetrică (deoarece $\langle\psi|\Phi\rangle = \langle\Phi|\psi\rangle$). Mai mult, $\langle\psi|\psi\rangle$ este întotdeauna un număr real nenegativ, astfel încât are o rădăcină pătrată nenegativă pe care o vom numi *lungimea* (sau mărimea) lui $|\psi\rangle$.

Deoarece putem înmulţi orice vector de stare cu un număr complex diferit de zero fără să-i schimbăm interpretarea fizică, putem întotdeauna *normaliza* starea, astfel încât să aibă lungimea unitate, realizând astfel un *vector unitate* sau o *stare normalizată*. Rămâne totuşi ambiguitatea că starea vectoare poate fi multiplicată cu un factor de fază (un număr de forma $e^{i\theta}$, cu θ real, conform paragrafului 5.10).

5.13.Descrierea lui R cu ajutorul spaţiului Hilbert

Cum putem reprezenta operaţia **R** în termenii spaţiului Hilbert? Să considerăm cel mai simplu caz al unei măsurători de tip 'da/nu', unde anumite piese ale aparatului înregistrează DA atunci când arată că obiectul cuantic măsurat are o anumită proprietate şi NU dacă nu poate înregistra acest lucru (sau, în mod echivalent, dacă înregistrează că obiectul cuantic nu are proprietatea respectivă). O astfel de măsurătoare da/nu include şi posibilitatea ca alternativele NU să fie o măsurătoare nulă şi acesta este pentru noi cazul cel mai interesant. De exemplu,

unii dintre detectorii de fotoni din paragraful 5.8 execută măsurători de acest tip. Ei vor înregistra DA dacă fotonul este capturat şi NU în caz contrar. În acest caz, măsurătoarea NU este într-adevăr o măsurătoare număr — ceea ce nu o face să fie mai puţin măsurătoare, provocând 'saltul' stării cuantice iniţiale într-o stare ortogonală cu cea care ar fi dat un răspuns DA. De asemenea, măsurătorile de spin Stern-Gerlach din paragraful 5.10 pentru un atom de spin ½ sunt de asemenea măsurători de acest fel; putem spune că răspunsul este DA dacă spinul atomului este găsit a fi $|\uparrow>$, ceea ce se întâmplă dacă atomul este găsit în raza care corespunde lui $|\uparrow>$ şi NU dacă nu este găsit în acea rază, ceea ce înseamnă că atomul se află într-o stare ortogonală pe $|\uparrow>$, adică $|\downarrow>$

Măsurătorile mai complicate pot fi întotdeauna considerate o succesiune de măsurători da/nu. Să considerăm un atom de spin $\frac{1}{2}n$, de exemplu. Pentru a obţine $n+1$ posibilităţi diferite pentru o măsurătoare a cantităţii de spin în direcţia 'sus', putem întreba, mai întâi, dacă starea de spin este $|\uparrow\uparrow\uparrow\ \ \uparrow>$ Facem acest lucru încercând să detectăm atomul în raza corespunzătoare acestei stări de spin 'sus' Dacă obţinem DA, atunci am terminat; dacă obţinem NU, atunci avem de-a face cu o măsurătoare nulă şi putem trece mai departe întrebând dacă starea de spin este $|\downarrow\uparrow\uparrow\ \ \uparrow>$ şi aşa mai departe. În fiecare caz, răspunsul NU va fi o măsurătoare nulă, indicând pur şi simplu că răspunsul DA nu a fost încă obţinut. Mai detaliat, să spunem că starea iniţială este;

$$z_0|\uparrow\uparrow\uparrow\ \ \uparrow> + z_1|\downarrow\uparrow\uparrow\ \ \uparrow> + z_2|\downarrow\downarrow\uparrow\ \ \uparrow> +\ldots + z_n|\downarrow\downarrow\downarrow\ldots\downarrow>$$

şi că ne întrebăm dacă spinul este 'sus' Dacă găsim răspunsul DA, atunci putem fie să spunem că starea este într-adevăr $|\uparrow\uparrow\uparrow\ \ \uparrow>$, sau putem să considerăm că starea 'sare' la $|\uparrow\uparrow\uparrow\ \ \uparrow>$ Însă dacă răspunsul este NU, atunci, ca rezultat la măsurătorii nule, trebuie să considerăm că starea 'sare' în cea ortogonală:

$$z_1|\downarrow\uparrow\uparrow\ \ \uparrow> + z_2|\downarrow\downarrow\uparrow\ \ \uparrow> +\ldots + z_n|\downarrow\downarrow\downarrow\ldots\downarrow>$$

şi vom încerca din nou, întrebând dacă starea este $|\downarrow\uparrow\uparrow\ \ \uparrow>$ Dacă obţinem de această dată DA, atunci vom spune că starea este într-adevăr $|\downarrow\uparrow\uparrow\ \ \uparrow>$, sau, altfel spus, că starea a 'sărit' la $|\downarrow\uparrow\uparrow\ \ \uparrow>$ Dacă răspunsul este însă NU, atunci starea 'sare' la

$$z_2|\downarrow\downarrow\uparrow\ \ \uparrow> +\ldots + z_n|\downarrow\downarrow\downarrow\ \ \downarrow>$$

şi aşa mai departe.

Această 'săritură' pe care vectorul de stare o comite — sau, cel puţin, *pare* să o comită — este cel mai neliniştitor aspect al teoriei cuantice. Ar fi

probabil cinstit să începem prin a spune că, cei mai mulţi dintre fizicieni, fie găsesc că acest proces este greu de acceptat ca o trăsătură a realităţii fizice, fie refuză cu totul să accepte că realitatea se poate comporta în acest mod absurd. Cu toate acestea, indiferent ce punct de vedere adoptăm privind 'realitatea' implicată, avem de-a face cu o trăsătură esenţială a formalismului cuantic.

În descrierile de mai înainte, m-am bazat pe ceea ce este numit uneori *postulatul proiecţiei*, care specifică forma pe care 'săritura' trebuie să o ia (de exemplu $z_0|\uparrow\uparrow...\uparrow> + z_1|\downarrow\uparrow...\uparrow> + \quad + z_n|\downarrow\downarrow...\downarrow>$ sare la $z_1|\downarrow\uparrow...\uparrow> + \quad + z_n|\downarrow\downarrow...\downarrow>$). Vom vedea în curând motivul geometric pentru care folosesc această terminologie. Anumiţi fizicieni pretind că postulatul proiecţiei este o presupunere ne-necesară în teoria cuantică. În loc să vorbească despre măsurători nule, ei se referă, în general, la măsurători în care starea cuantică este *perturbată* de anumite interacţiuni fizice. O astfel de perturbaţie apare atunci când, în exemplele de mai sus, se obţine răspunsul DA de exemplu atunci când un detector de fotoni absoarbe fotonul pentru a-l înregistra, sau atunci când un atom, după ce trece prin aparatul Stern-Gerlach, este găsit într-unul din fasciculele particulare (fasciculul DA, de exemplu). Pentru o măsurătoare nulă de tipul celei pe care o consider aici, (răspunsul NU) postulatul proiecţiei este într-adevăr esenţial, pentru că fără el nu putem stabili ce anume prezice teoria că se va întâmpla în următoarele experimente.

Pentru a fi mai explicit în legătură cu ce anume spune postulatul proiecţiei, să vedem mai întâi cum arată descrierea cu ajutorul spaţiului Hilbert. Voi considera un tip particular de măsurătoare — pe care o voi numi măsurătoare *primitivă* — în care măsurătoarea este de tipul da/nu, însă răspunsul DA stabileşte că starea cuantică este — sau a 'sărit' la — o stare particulară $|\alpha>$ (sau un multiplu nenegativ $u|\alpha>$). Astfel, pentru o măsurătoare primitivă, răspunsul DA determină starea fizică ca pe *un* lucru particular, pe când pentru răspunsul NU pot exista mai multe alternative. Măsurătorile de spin de mai sus, în care încercăm să stabilim dacă spinul este într-o stare particulară, ca de exemplu $|\downarrow\downarrow\uparrow\uparrow \quad \uparrow>$, sunt exemple de măsurători primitive.

Răspunsul NU pentru o măsurătoare primitivă *proiectează* starea în ceva ortogonal pe $|\alpha>$. Puteţi vedea imaginea geometrică a acestei proiecţii în figura 5.24. Starea originală este $|\psi>$, reprezentată prin săgeata mare şi, ca rezultat al măsurătorii, ea fie 'sare' la un multiplu de $|\alpha>$, atunci când răspunsul este DA,

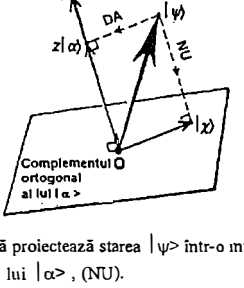


Fig. 5.24. O măsurătoare primitivă proiectează starea $|\psi\rangle$ într-o multitudine de stări $|\alpha\rangle$ alese (DA), sau în complementul ortogonal al lui $|\alpha\rangle$, (NU).

fie este proiectată pe spaţiul ortogonal cu $|\alpha\rangle$, dacă răspunsul este NU. În cazul lui NU, nu avem nici un fel de problemă; în cazul teoriei cuantice standard trebuie să considerăm că exact asta se întâmplă cu starea cuantică. În cazul lui DA situaţia este complicată de faptul că sistemul cuantic a interacţionat cu aparatul de măsură, iar starea a devenit astfel ceva mai complicat decât $|\alpha\rangle$. De fapt, starea evoluează în ceea ce se numeşte o *stare legată* care împleteşte stările sistemului cuantic şi ale aparatului de măsură (stările legate vor fi discutate în paragraful 5.17). Dar evoluţia stării cuantice continuă *ca şi când* ar 'sări' într-adevăr într-un multiplu al lui $|\alpha\rangle$, astfel încât evoluţia ulterioară să poată continua lipsită de ambiguităţi.

Putem exprima această 'săritură' algebric, în următorul mod. Vectorul de stare $|\psi\rangle$ poate fi întotdeauna scris (în mod unic, dacă $|\alpha\rangle$ este dat) ca:

$$|\psi\rangle = z|\alpha\rangle + |\chi\rangle ,$$

unde $|\chi\rangle$ este ortogonal pe $|\alpha\rangle$. Vectorul $z|\alpha\rangle$ este proiecţia ortogonală a lui $|\psi\rangle$ pe raza determinată de $|\alpha\rangle$, iar $|\chi\rangle$ este proiecţia ortogonală a lui $|\psi\rangle$ pe spaţiul *ortogonal complemetar* lui $|\alpha\rangle$, (adică pe spaţiul tuturor vectorilor otogonali pe $|\alpha\rangle$). Dacă rezultatul măsurătorii este DA, atunci trebuie să considerăm că vectorul de stare a 'sărit' la $z|\alpha\rangle$ (sau, mai simplu, la $|\alpha\rangle$) ca punct de plecare al evoluţiei lui ulterioare; dacă este NU, atunci 'sare' la $|\chi\rangle$.

Cum obţinem probabilităţile care trebuie atribuite acestor rezultate alternative? Pentru a putea folosi 'regula pătratului modulului' pe care am întâlnit-o mai înainte, cerem acum ca $|\alpha\rangle$ să fie un vector unitate şi alegem un vector unitate $|\Phi\rangle$ în direcţia lui $|\chi\rangle$, astfel încât $|\chi\rangle = w|\Phi\rangle$. Avem acum $|\psi\rangle = z|\alpha\rangle + w|\Phi\rangle$

unde, de fapt, $z = \langle\alpha|\psi\rangle$ şi $w = \langle\Phi|\psi\rangle$ şi unde putem obţine probabilităţile relative de DA şi NU ca raport între $|z|^2$ şi $|w|^2$. Dacă $|\psi\rangle$ este un vector unitate, atunci $|z|^2$ şi $|w|^2$ sunt probabilităţile *actuale* ale lui DA şi,respectiv, NU

Mai există şi alt mod de a formula aceasta, într-un chip ceva mai simplu în prezentul context (şi las asta ca exerciţiu cititorului interesat, pentru a verifica că este vorba cu adevărat de echivalenţă). Pentru a obţine probabilitatea curentă a fiecărui răspuns DA sau NU examinăm pătratul lungimii vectorului $|\psi\rangle$ (fără a presupune că a fost normalizat în vector unitate) şi urmărim în ce proporţie această mărime se reduce în fiecare din proiecţiile respective. Factorul de reducere, în fiecare caz, este probabilitatea cerută.

În sfârşit, trebuie menţionat că, pentru o măsurătoare *generală* da/nu (acum nu neapărat o măsurătoare primitivă), în care stările DA nu aparţin neapărat unei singure raze, discuţia decurge la fel. Vom avea un subspaţiu **D** pentru DA, precum şi un subspaţiu **N** pentru NU. Aceste subspaţii sunt ortogonal complementare în sensul că fiecare vector al unuia este ortogonal cu toţi vectorii celuilalt, ele acoperind întregul spaţiu Hilbert original. Postulatul proiecţiei afirmă că, la măsurare, vectorul original $|\psi\rangle$ ajunge proiectat ortogonal în **D** când se obţine un răspuns DA, respectiv în **N**, pentru un răspuns NU. Din nou, probabilităţile respective sunt date prin factorii de reducere în proiecţii a pătratului lungimii vectorului de stare (vezi **MN**, p.286, figura 6.23).

Condiţia din postulatul proiecţiei este aici mai puţin clară decât în cazul măsurătorii nule de mai înainte, deoarece, cu o măsurătoare afirmativă, starea rezultat devine legată cu starea aparatelor de măsură. Pentru aceste considerente, în cele ce urmează, mă voi limita la măsurătorile simple *primitive*, în care spaţiul DA constă dintr-o singură rază (multiplicată cu $|\psi\rangle$). Este suficient pentru demersul de faţă.

5.14. Măsurători comutative

În general, dacă avem de-a face, în cazul unui sistem cuantic, cu măsurători succesive, iar ordinea în care aceste măsurători sunt efectuate poate fi importantă. Acelea pentru care ordinea este relevantă se numesc *măsurători necomutative.*

Dacă ordinea măsurătorilor nu joacă nici un rol (dacă nu exista nici măcar un factor de fază ca diferență) atunci spunem că măsurătorile *comută*. În termenii spațiului Hilbert, faptul că o măsurătoare este neomutativă revine la a spune că proiecțiile ortogonale succesive ale unei stări date ne vor da un rezultat diferit în funcție de ordinea în care sunt făcute. Pentru măsurătorile comutative, această ordine nu provoacă nici o diferență în rezultatul final.

Ce se întâmplă cu măsurătorile primitive? Nu este greu de văzut că, pentru ca o pereche de măsurători primitive distincte să comute, condiția este ca raza DA a unei măsurători să fie ortogonală pe raza DA a celeilalte măsurători.

De exemplu, cu măsurătorile primitive de spin, considerate în paragraful 5.10, care trebuiau efectuate asupra unui atom de spin ½*n*, ordinea este într-adevăr irelevantă, deoarece diferitele stări luate în considerare, adică $|\uparrow\uparrow...\uparrow>$, $|\downarrow\uparrow\uparrow...\uparrow>$,..., $|\downarrow\downarrow...\downarrow>$ sunt toate ortogonale una pe alta. Astfel, ordinea particulară pentru aceste măsurători primitive nu joacă nici un rol în rezultatul final și măsurătorile comută. Totuși, nu acesta este cazul general dacă diferitele măsurători ale spinului au fost efectuate în diferite direcții. În general, ele nu comută.

5.15. „și"-ul mecanicii cuantice

Există o procedură standard în mecanica cuantică pentru tratarea sistemelor care conțin mai mult decât o parte independentă. Această procedură va fi necesară, în particular, pentru discutarea sistemelor alcătuite din două particule separate de spin 3/2 (vezi paragraful 5.18), de felul celor pe care Tinichelele Chintesențiale le plasaseră în centrul celor două dodecaedre magice din paragraful 5.3. Același tip de procedură este necesară pentru descrierea cuantică a unui detector în stare legată cu particula pe care o detectează.

Să considerăm mai întâi că avem un sistem alcătuit din două părți independente (care nu interacționează). Vom presupune că fiecare dintre ele, în absența celeilalte, ar fi descrisă de un vector de stare $|\alpha>$, respectiv $|\beta>$. Cum trebuie să descriem acum sistemul în care *ambele* particule sunt prezente? Procedura normală este să formăm ceea ce se numește *produsul tensorial* (sau produsul *extern*) al acestor vectori, scris $|\alpha>|\beta>$

Ne putem imagina că acest produs ne furnizează metoda standard în mecanica cuantică de reprezentare a noţiunii 'şi', în sensul în care cele două sisteme cuantice independente reprezentate, respectiv prin $|\alpha>$ şi $|\beta>$ sunt ambele prezente în acelaşi timp (de exemplu, $|\alpha>$ poate fi reprezentată printr-un electron localizat în A iar $|\beta>$ poate fi reprezentată de un atom de hidrogen aflat undeva la mare distanţă în B. Starea, în care electronul este în A şi atomul de hidrogen este în B, va fi reprezentată prin $|\alpha>|\beta>$). Cantitatea $|\alpha>|\beta>$ este un *singur* vector de stare, să-i spunem $|\chi>$ astfel încât putem scrie $|\chi>=|\alpha>|\beta>$, de exemplu.

Trebuie să subliniem că acest concept, 'şi', este complet diferit de superpoziţia liniară cuantică, care ar avea descrierea $|\alpha>+|\beta>$ sau, mai general, $z|\alpha>+w|\beta>$, unde z şi w sunt factori de pondere complecşi. Dacă, de exemplu, $|\alpha>$ şi $|\beta>$ sunt stări posibile ale unui foton individual, localizat, să spunem, în A şi, în acelaşi timp cu posibilitatea de a fi localizat în B, atunci, $|\alpha>+|\beta>$ este de asemenea o stare posibilă pentru un *singur* foton a cărui poziţie este divizată între A şi B, conform predicţiilor stranii ale teoriei cuantice — şi nu o descriere a *doi* fotoni. O *pereche* de fotoni, unul în A şi altul în B, vor fi reprezentaţi prin starea $|\alpha>|\beta>$

Produsul tensorial satisface tipul de reguli algebrice pe care l-am aştepta de la un 'produs':

$$(z|\alpha>)|\beta> = z(|\alpha>|\beta>) = |\alpha>(z|\beta>),$$
$$(|\alpha>+|\gamma>)|\beta> = |\alpha>|\beta>+|\gamma>|\beta>,$$
$$|\alpha>(|\beta>+|\gamma>) = |\alpha>|\beta>+|\alpha>|\gamma>,$$
$$(|\alpha>|\beta>)|\gamma> = |\alpha>(|\beta>|\gamma>),$$

cu excepţia faptului că nu este întru totul corect să scriem $|\alpha>|\beta> = |\beta>|\alpha>$ Cu toate acestea, ar fi nerezonabil să ne gândim că putem interpreta cuvântul 'şi' în contextul mecanicii cuantice ca o implicaţie a faptului că sistemul combinat '$|\alpha>$ şi $|\beta>$' este diferit fizic de sistemul '$|\beta>$ şi $|\alpha>$' Putem evita această dificultate dacă analizăm cu mai mare atenţie felul în care Natura se comportă. În loc să interpretăm starea $|\alpha>|\beta>$, aşa cum o fac matematicienii, ca pe un produs tensorial, putem considera că ea presupune ceea ce fizicienii numesc un *produs Grassmann*. Avem atunci o regulă adiţională:

$$|\beta>|\alpha>=\pm|\alpha>|\beta>,$$

unde semnul minus apare atunci când *ambele* stări $|\alpha>$ şi $|\beta>$ au un număr *impar* de particule cu spin neîntreg (spinul unor asemenea particule are una din valorile ½, 3/2, 5/2, 7/2,.... — astfel de particule se numes *fermioni*; particulele de spin 0,1,2.... se numesc *bosoni*, ele nu contribuie la semnul expresiei precedente). Din nou nu este nevoie să vă preocupaţi de aspectele tehnice. În măsura în care starea fizică este cea în cauză în descrierea noastră, '$|\alpha>$ şi $|\beta>$' este acelaşi lucru cu '$|\beta>$ şi $|\alpha>$'

Pentru stările care au trei sau mai multe părţi independente, nu trebuie decât să repetăm acest procedeu. Astfel, cu trei părţi, ale căror stări individuale sunt $|\alpha>$, $|\beta>$ şi $|\gamma>$, starea în care toate cele trei părţi sunt prezente în acelaşi timp poate fi scrisă ca $|\alpha>|\beta>|\gamma>$, acelaşi lucru ca mai sus (interpretat în termenii produsului Grassmann) adică $(|\alpha>|\beta>)|\gamma>$ sau, echivalent, $|\alpha>(|\beta>|\gamma>)$. Cazul în care avem patru sau mai multe părţi este similar.

O proprietate importantă a evoluţiei **U**, pentru sistemele $|\alpha>$ şi $|\beta>$ care nu interacţionează unul cu altul este că evoluţia sistemului combinat este chiar evoluţia combinată a sistemelor individuale. Astfel, dacă după un anumit timp t, sistemul $|\alpha>$ va evolua (de unul singur) în $|\alpha'>$, iar $|\beta>$ în $|\beta'>$, atunci sistemul combinat $|\alpha>|\beta>$ va evolua în acelaşi timp t, în $|\alpha'>|\beta'>$ Mai mult, rezultă că, dacă avem trei sisteme neinteractive $|\alpha>$, $|\beta>$ şi $|\gamma>$ ale unui sistem $|\alpha>|\beta>|\gamma>$, iar părţile evoluează, respectiv, în $|\alpha'>$, $|\beta'>$ şi $|\gamma'>$, atunci, sistemul combinat va evolua de asemnea în $|\alpha'>|\beta'>|\gamma'>$ Acelaşi lucru este valabil pentru patru sau mai multe părţi.

Să observăm că acest lucru este foarte asemănător cu proprietatea de *liniaritate* a lui **U** la care ne-am referit în paragraful 5.7, conform căreia stările suprapuse evoluează ca superpoziţia evoluţiilor stărilor individuale. De exemplu, $|\alpha>+|\beta>$ va evolua în $|\alpha'>+|\beta'>$ Totuşi, este important să realizăm că avem de-a face cu un lucru foarte diferit. Nu este nimic surprinzător în faptul că un sistem compus din părţi individuale care nu interacţionează între ele va evolua, ca întreg, ca şi cum fiecare parte separată n-ar şti de prezenţa celorlalte. Este evident esenţial ca părţile să nu interacţioneze, altfel proprietatea ar fi falsă. Proprietatea de liniaritate, pe de altă parte, este o reală surpriză. În acest caz, conform **U**, sistemele aflate în stare de superpoziţie evoluează aproape independent, de parcă n-ar exista între ele nici un fel de interacţiune. Doar acest fapt este suficient pentru a ne pune întrebarea asupra adevărului proprietăţii de liniaritate —

care însă este solid confirmată experimental pentru fenomene cuantice. Doar operaţia **R** pare s-o violeze. Dar ne vom întoarce la acestea ceva mai târziu.

5.16. Ortogonalitatea stărilor produs

Există o anumită ciudăţenie a stărilor produs, pe care le-am descris, în relaţie cu noţiunea de ortogonalitate. Dacă avem două *stări ortogonale* $|\alpha\rangle$ şi $|\beta\rangle$, atunci putem aştepta ca stările $|\psi\rangle|\alpha\rangle$ şi $|\psi\rangle|\beta\rangle$ să fie de asemenea ortogonale, pentru orice $|\psi\rangle$. De exemplu, $|\alpha\rangle$ şi $|\beta\rangle$ pot fi două stări alternative posibile ale unui foton: starea $|\alpha\rangle$ ar putea fi cea detectată de o celulă fotoelectrică, pe când starea $|\beta\rangle$ va fi starea *inferată* a fotonului atunci când fotocelula eşuează în înregistrarea lui (măsurătoare nulă).

Am putea considera acum că fotonul este doar o parte a unui sistem *combinat*, căruia îi adăugăm pur şi simplu alte obiecte alt foton, de pildă, aflat undeva pe Lună starea acestui al doilea obiect fiind $|\psi\rangle$. În acest caz, avem, pentru sistemul combinat, cele două stări alternative $|\psi\rangle|\alpha\rangle$ şi $|\psi\rangle|\beta\rangle$.

Simpla includere a stării $|\psi\rangle$ în descriere nu va afecta cu siguranţă ortogonalitatea acestor două stări. Într-adevăr, cu definiţia normală a stărilor produs (cea folosind 'produsul tensorial' ca opusă tipului de produs Grassmann folosit aici) şi acesta va fi cazul aici, iar ortogonalitatea lui $|\psi\rangle|\alpha\rangle$ şi $|\psi\rangle|\beta\rangle$ va rezulta din ortogonalitatea lui $|\alpha\rangle$ şi $|\beta\rangle$.

Totuşi, felul în care Natura pare să se comporte, în acord cu procedurile complete ale teoriei cuantice, nu este chiar atât de direct. Dacă starea $|\psi\rangle$ poate fi considerată total independentă atât de $|\alpha\rangle$ cât şi de $|\beta\rangle$, atunci prezenţa ei va fi irelevantă. Însă, tehnic vorbind, chiar starea unui foton de pe Lună nu poate fi privită ca fiind complet separată de cea implicată în procesul nostru de detecţie[4] (lucrul acesta are de-a face cu folosirea produsului de tip Grassmann în notaţia '$|\psi\rangle|\alpha\rangle$' folosită aici — mai familiar vorbind, are de-a face cu statistica Bose,

[4] În mod curios, acest tip de fenomen poate avea o profundă relevanţă pentru observaţiile noastre actuale. Efectul datorat lui Hanbury Brown şi Twiss (1954, 1956), conform căruia au fost măsurate diametrele anumitor stele, depinde de proprietatea 'bosonică' care leagă fotonii ce ating pământul proveniţi din părţi diferite ale stelei!

implicată în stările fotonice sau cu statistica Fermi, pentru electron, proton şi alte stări fermionice; vezi **MN** p. 327 şi Dirac, 1947).

Dacă am vrea să fim absolut riguroşi în aplicarea regulilor teoriei, ar trebui să considerăm toţi fotonii din Univers atunci când discutăm starea unui singur foton. Totuşi, chiar pentru un grad de acurateţe foarte mare, acest lucru nu este (din fericire) necesar. Vom considera pur şi simplu că, pentru orice stare $|\psi\rangle$ care în mod evident nu are nimic de-a face cu problema în discuţie, atunci cînd problema aceasta priveşte direct stările ortogonale $|\alpha\rangle$ şi $|\beta\rangle$, stările $|\psi\rangle|\alpha\rangle$ şi $|\psi\rangle|\beta\rangle$ vor fi într-adevăr ortogonale (chiar pentru un produs tensorial de tip Grassmann) cu un foarte mare grad de acurateţe.

5.17. Nonlocalizarea cuantică

Trebuie acum să ajungem să înţelegem fizica ce guvernează *efectele* EPR misterele **Z** cuantice exemplificate de dodecaedrele magice din 5.3, 5.4. Trebuie apoi să facem cunoştinţă cu misterele **X** fundamentale ale teoriei cuantice relaţiile paradoxale între cele două procese **U** şi **R** care se află în spatele *problemei măsurării* pe care o vom discuta în capitolul următor. Pentru amândouă aceste probleme va fi necesar să introducem aici încă o idee importantă, cea de *stări nelocalizate*.

Să vedem mai întâi ce anume este implicat într-un simplu proces de măsurare. Să considerăm o situaţie în care un foton se află într-o stare suprapusă, de exemplu, $|\alpha\rangle+|\beta\rangle$ unde starea $|\alpha\rangle$ ar activa detectorul, pe când starea $|\beta\rangle$, ortogonală pe $|\alpha\rangle$, ar lăsa detectorul neperturbat (un exemplu de acest tip a fost considerat în paragraful 5.8, când detectorul G întâlnea starea $-|F\rangle-|G\rangle$; aici, $|G\rangle$ va activa detectorul, pe când $|F\rangle$ îl va lăsa neactivat). Voi presupune că detectorului însuşi i se poate atribui o stare cuantică, să-i spunem $|\Psi\rangle$. Asta este o practică uzuală în teoria cuantică. Nu-mi este în întregime clar dacă are sens să atribuim o descriere cuantică unui obiect clasic, însă în discuţii de acest fel, procedura este considerată normală. În orice caz, putem presupune că elementele detectorului, pe care fotonul le va întâlni *iniţial*, sunt într-adevăr lucruri care pot fi tratate cu ajutorul regulilor standard ale mecanicii cuantice. Cei care mai păstrează un dubiu, cu privire la şansa de a trata detectorul ca întreg după aceste

legi, pot considera că vectorul de stare $|\Psi>$ se referă la aceste elemente cuantice inițiale (atomi, particule, molecule).

Chiar înainte ca fotonul să atingă detectorul (sau, mai degrabă, chiar înainte ca porțiunea $|\alpha>$ a funcției de undă a fotonului să atingă detectorul), situația fizică constă în starea detectorului și starea fotonului, adică, $|\Psi>(|\alpha> + |\beta>)$ Vom găsi:

$$|\Psi>(|\alpha> + |\beta>) = |\Psi>|\alpha> + |\Psi>|\beta>$$

Aceasta este o superpoziție a stării $|\Psi>|\alpha>$ care descrie detectorul (elementele) și fotonul care se apropie, cu starea $|\Psi>|\beta>$, care descrie elementele detectorului și fotonul aflat în altă parte. Să presupunem că următorul lucru care se întâmplă, în acord cu evoluția **U**, este că starea $|\Psi>|\alpha>$ (detector și foton apropiindu-se de el) devine o stare nouă $|\Psi_Y>$ (care indică faptul că detectorul a înregistrat un răspuns DA) în virtutea interacțiunilor fotonului cu elementele detectorului pe care îl întâlnește. Vom presupune de asemenea că, dacă fotonul *nu* întâlnește detectorul, acțiunea lui **U** va provoca starea detectorului $|\Psi>$ să evolueze (prin ea însăși) în $|\Psi_N>$ (detectorul înregistrând răspunsul NU) iar $|\beta>$ va evolua în $|\beta'>$. Atunci, din proprietățile evoluției Schrödinger pe care le-am văzut în secțiunea precedentă, starea totală va deveni:

$$|\Psi_Y> + |\Psi_N>|\beta'>.$$

Acesta este un exemplu particular de *stare nonlocalizată*, unde 'nonlocalizarea' vine din faptul că nu putem scrie starea totală ca un produs de stări pentru unul dintre subsisteme (fotonul) cu celălalt subsistem (detectorul). De fapt, starea $|\Psi_Y>$ însăși este foarte probabil să fie o stare legată (nonlocalizată) de întregul mediu în care se află, însă aceasta depinde de detaliile unor interacțiuni ulterioare, care nu ne interesează aici.

Observăm că stările $|\Psi>|\alpha>$ și $|\Psi>|\beta>$, a căror superpoziție reprezintă starea sistemului combinat chiar înainte de interacție, sunt (esențial) *ortogonale* -- deoarece $|\alpha>$ și $|\beta>$ sunt ele însele ortogonale, iar $|\Psi>$ este total independentă de oricare din ele. Astfel, stările $|\Psi_Y>$ și $|\Psi_N>|\beta'>$ în care evoluează sub acțiunea lui **U**, trebuie la rândul lor să fie ortogonale (**U** păstrează întotdeauna ortogonalitatea.) Starea $|\Psi_Y>$ poate evolua mai departe în ceva ce este macroscopic observabil, ca, de exemplu, un zgomot (*click*), indicând că fotonul a fost într-adevăr detectat. Dacă nu se aude nici un astfel de click, înseamnă că starea trebuie să fi fost -- adică trebuie să fi 'sărit' la posibilitatea

ortogonală $|\Psi_N\rangle|\beta'\rangle$. Dacă nu auzim nici un click, atunci posibilitatea contrafactuală ca el să se fi produs (fără a se produce în mod actual) va provoca 'saltul' stării inițiale la $|\Psi_N\rangle|\beta'\rangle$ care nu este o stare nonlocalizată. Măsurătoarea nulă a *dezlegat* (localizat) starea.

O trăsătură caracteristică a stărilor legate este că 'saltul' ce apare cu operația R are un caracter aparent nonlocal (sau chiar aparent retroactiv) — acțiune care este chiar mai enigmatică decât o simplă măsurătoare nulă. O astfel de nonlocalitate apare în particular le ceea ce numim *efecte* EPR (Einstein, Podolski, Rosen) — enigme cuantice veritabile care se află printre cele mai surprinzătoare dintre misterele Z. Ideea acestor 'efecte' i-a venit prima oară lui Einstein, într-o încercare de a arăta că formalismul teoriei cuantice nu poate furniza o descriere completă a Naturii. De atunci au fost enunțate multe variante ale aceluiași fenomen (ca, de exemplu, dodecaedrele magice din paragraful 5.3), iar unele variante au fost confirmate experimental (direct) ca trăsături *actuale* ale lumii în care trăim (vezi paragraful 5.4).

Efectele EPR apar în următorul tip de situație. Considerați o stare inițială cunoscută $|\Omega\rangle$ a unui sistem fizic, care evoluează (prin U) într-o superpoziție de două stări ortogonale, fiecare dintre ele fiind produsul unei perechi de stări independente care descriu o pereche de părți (particule) separate spațial, astfel încât $|\Omega\rangle$ evoluează în starea legată:

$$|\psi\rangle|\alpha\rangle + |\phi\rangle|\beta\rangle$$

Vom presupune că $|\psi\rangle$ și $|\phi\rangle$ sunt posibilități ortogonale pentru una dintre părți, iar $|\alpha\rangle$ și $|\beta\rangle$ sunt posibilități ortogonale pentru cealaltă parte. O măsurătoare care ne spune dacă prima parte este în starea $|\psi\rangle$ sau $|\phi\rangle$ va determina instantaneu dacă a doua parte este în starea $|\alpha\rangle$ sau $|\beta\rangle$

Până aici, nimic misterios. Am putea susține că avem aceeași situație ca în cazul șosetelor lui Bertlmann (paragraful 5.4). În cunoașterea faptului că aceste două șosete trebuie să difere la culoare — și să presupunem că știm că astăzi el a ales să poarte una verde și una roz — observația că șoseta stângă este verde (starea $|\psi\rangle$) sau roz (starea $|\phi\rangle$) va determina instantaneu faptul că șoseta dreaptă este în mod corespunzător fie roz (starea $|\alpha\rangle$) fie verde (starea $|\beta\rangle$). Totuși, efectele nonlocalizării cuantice pot diferi profund de acest caz și nici o explicație de tipul șosetelor lui Bertlmann nu poate da seama de efectele observaționale. Problema apare atunci când trebuie efectuate tipuri *alternative* de

măsurătoare asupra celor două părţi ale sistemului.

Iată un exemplu. Să presupunem că starea iniţială $|\Omega_0>$ descrie starea de spin a unei particule ca fiind de spin 0. Această particulă se dezintegrează în alte două particule, fiecare de spin ½, care se îndepărtează una de alta la distanţă foarte mare, spre dreapta şi spre stânga. Din proprietăţile momentului cinetic şi din conservarea sa rezultă că direcţiile spinului celor două particule trebuie să fie *opuse* una alteia şi, pentru starea totală de spin 0 în care $|\Omega_0>$ evoluează, găsim,

$$|\Omega> = |L\uparrow>|R\downarrow> - |L\downarrow>|R\uparrow>$$

unde 'L' se referă la particula din dreapta, iar 'R' la particula din stânga (iar semnul apare din convenţiile standard). Astfel, dacă am ales să măsurăm spinul particulei din dreapta în direcţia 'sus', răspunsul DA (adică găsim $|L\uparrow>$) va pune automat particula din partea dreaptă în starea $|R\downarrow>$ de spin 'jos' Răspunsul NU ($|L\downarrow>$) va pune automat particula din dreapta în starea de spin sus, ($|R\uparrow>$). Pare că măsurătoarea unei particule într-un loc poate afecta instantaneu starea unei particule aflată într-un loc diferit — dar până aici nu avem nimic mai misterios decât sunt şosetele lui Bertlmann!

Totuşi, starea noastră legată poate fi reprezentată şi altfel, corespunzător unui alt mod de a efectua măsurătoarea. De exemplu, putem alege să măsurăm spinul particulei din stânga într-o direcţie *orizontală*, astfel încât, acum, DA va corespunde, să spunem, lui $|L\rightarrow>$, iar NU va corespunde lui $|L\leftarrow>$ Găsim deci (printr-un calcul simplu, vezi MN) că *aceeaşi* stare combinată de mai înainte, poate fi scrisă într-un mod diferit ca $|\Omega> = |L\leftarrow>|R\rightarrow> - |L\rightarrow>|R\leftarrow>$ Astfel, găsim că răspunsul DA la stânga pune automat particula din dreapta în starea $|R\rightarrow>$, iar răspunsul NU o pune în starea $|R\leftarrow>$ Un proces asemănător se petrece *indiferent* ce direcţie se întâmplă să alegem pentru măsurătorile spinului particulei din stânga.

Ce este remarcabil în această situaţie este faptul că simpla *alegere* a direcţiei spinului particulei din stânga pare să *fixeze* direcţia spinului particulei din dreapta. De fapt, până când nu obţinem rezultatul măsurătorii din stânga, nu există nici un fel de informaţie privind particula din dreapta. Simpla 'fixare a direcţiei spinului' nu provoacă prin ea însăşi nimic observabil. Dar, în ciuda acestui fapt cunoscut, se întâmplă deseori ca unii să susţină că efectele EPR ar putea fi folosite pentru a trimite semnale dintr-un loc în altul *instantaneu*, din moment ce reducţia vectorului de stare 'reduce' starea cuantică a particulelor EPR simultan, indiferent

cât sunt ele de departe una de alta. În realitate, nu există nici o modalitate de a trimite, prin acest procedeu, un semnal de la particula din stânga la cea din dreapta (Ghirardi *et al.* 1980).

Folosirea standard a formalismului mecanicii cuantice furnizează într-adevăr următoarea imagine: în momentul în care o măsurătoare este efectuată asupra unei particule, să spunem cea din stânga, întreaga stare se reduce instantaneu şi, în locul celei nelocalizate în care nici una dintre particule nu are o stare 'individuală' bine definită, obţinem o stare în care cele două particule sunt 'dezlegate' şi au stări de spin individuale bine definite. În descrierea matematică a vectorului de stare, măsurătoarea efectuată asupra particulei din stânga are un efect instantaneu asupra celei din dreapta. Însă, aşa cum am mai spus, acest 'efect instantaneu' nu este de un tip care să ne permită să trimitem un semnal fizic.

Conform principiilor relativităţii, semnalele fizice — obiecte capabile să transmită informaţie — sunt obligate să se propage cu viteze mai mici sau egale cu viteza luminii. Efectele EPR, însă, nu pot fi privite în acest fel. Nu ar fi consistent cu predicţiile teoriei cuantice să le tratăm ca semnale finite, propagându-se cu viteza luminii. (Exemplul dodecaedrelor magice ilustrează exact acest fapt; 'legătura' dintre dodecaedrul meu şi cel al colegului meu are efecte imediate — şi nu întârziate cu 4 ani, timpul necesar luminii să se propage de la unul la altul. Vezi paragrafele 5.3, 5.4 şi nota 4.) Deci, efectele EPR nu pot fi semnale în sensul obişnuit.

După toate acestea, ne-am putea întreba cum anume au efectele EPR consecinţe observabile de orice fel. Că au astfel de consecinţe rezultă din celebra teoremă a lui John Bell (conform paragrafului 5.4). Probabilităţile cumulate pe care teoria cuantică le prezice pentru diferitele măsurători posibile care pot fi efectuate asupra particulelor de spin ½ (cu alegerea independentă a direcţiilor de spin asupra particulelor din stânga şi din dreapta) nu pot fi obţinute din nici un model clasic de obiecte independente care nu comunică între ele (vezi câteva exemple de acest tip în MN). Dodecaedrele magice din paragraful 5.3 ne dau efecte chiar mai puternice, unde misterul apare generat de constrângeri precise de tip DA/NU în loc de probabilităţi. Astfel, chiar dacă particulele din dreapta şi din stânga nu comunică una cu alta, în sensul că nu sunt în stare să trimită mesaje de la una la alta, ele nu sunt mai puţin legate una de alta — nu pot fi considerate obiecte independente — până când sunt *dezlegate* de către măsurătoare. Nelocalizarea cuantică este un

fenomen misterios cuprins undeva între comunicare şi separaţie completă — şi nu are nici un fel de analog clasic. Mai mult, e un efect care nu descreşte cu distanţa (spre deosebire de, să zicem, legea atracţiei electrice sau gravitaţionale). Einstein găsea perspectiva unui asemenea efect profund neliniştitoare şi se referea la el ca la 'această fantomatică acţiune la distanţă' (vezi Mermin, 1985).

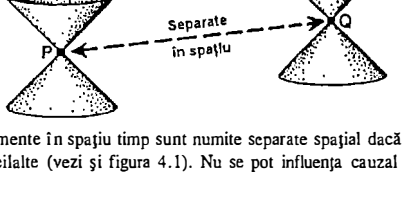


Fig. 5.25. Două evenimente în spaţiu timp sunt numite separate spaţial dacă fiecare se află în afara conului luminos al celeilalte (vezi şi figura 4.1). Nu se pot influenţa cauzal unul pe altul.

De fapt, nonlocalizarea cuantică pare să afecteze nu numai separarea spaţială ci şi separarea temporală. Dacă măsurătoarea este efectuată asupra unei componente a perechii EPR *înaintea* celeilalte componente, atunci prima măsurătoare este considerată, în descrierea cuantică obişnuită, cauza 'dezlegării', astfel încât a doua măsurătoare este o măsurătoare asupra unei particule individuale. Pe de altă parte, vom obţine exact aceleaşi consecinţe observabile dacă vom considera că *a doua măsurătoare* este cea care afectează (retroactiv) nonlocalizarea. Un alt mod de a exprima irelevanţa ordinii temporale asupra celor două măsurători este să spunem că ele *comută* (vezi paragraful 5.14).

Acest fel de simetrie este o trăsătură necesară a măsurătorilor EPR, pentru ca ele să fie consistente cu consecinţele observaţionale ale relativităţii restrânse. Măsurătorile operate asupra unor evenimente separate spaţial (adică evenimente care se găsesc în afara conurilor luminoase respective — vezi figura 5.25 şi discuţia din paragraful 4.4) *trebuie să comute* şi deci este nerelevantă ordinea în care sunt efectuate. Pentru a vedea că aşa este, putem considera întreaga situaţie ca fiind descrisă din punctul de vedere a doi observatori, în sisteme de referinţă deosebite, aşa cum se arată în figura 5.26 (vezi şi MN). Aceşti doi 'observatori' nu trebuie să aibă nici o relaţie cu cei care efectuează de fapt măsurătorile.) În situaţia reprezentată, cei doi observatori vor avea percepţii diferite asupra ordinii măsurătorilor. Deci, în măsurătorile de tip EPR, fenomenul de nonlocalizare

cuantică[5] nu ia în seamă nici separarea spațială, nici ordinea temporală!

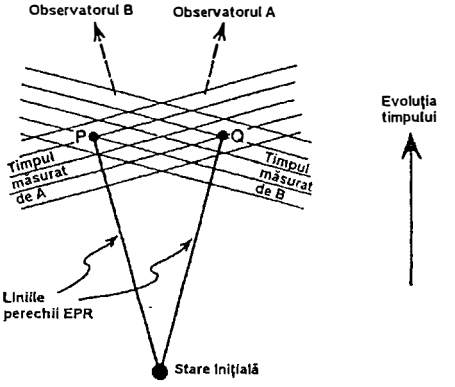


Fig. 5.26. Conform relativității restrânse, observatorii A și B, în mișcare relativă, au percepții diferite asupra ordinii evenimentelor separate spațial, P și Q (A îl crede primul pe Q, iar B, pe P).

5.18. Explicația dodecaedrelor magice

Pentru o pereche de particule EPR de spin ½, nonlocalizarea spațială sau temporală apare doar în probabilități. Însă nonlocalizarea cuantică este un fenomen mult mai concret și precis decât ceva de natură să influențeze doar probabilitățile.

Exemple ca cel al dodecaedrului magic (sau alte configurații de acest tip[10]) arată că strania nonlocalizare nu este doar o problemă de probabilitate, ci furnizează de asemenea efecte da/nu precise care nu pot fi explicate de fizica clasică.

Să vedem acum care este explicația cuantică pentru dodecaedrele magice din paragraful 5.3. Amintiți-vă că Tinichelele Chintesențiale au aranjat, acolo pe

[5] Pot fi date exemple în care proprietatea de nonlocalizare pentru o pereche de particule poate fi ea însăși o proprietate nelocală (vezi Zeilinger et al, 1992).

Betlegeuse, ca un sistem de spin total 0 să se împartă în doi atomi de spin 3/2 şi fiecare dintre aceşti doi atomi a fost suspendat cu grijă în centrul fiecărui dodecaedru. Cele două dodecaedre sunt apoi trimise, unul mie, celălalt colegului meu de pe α-Centauri, astfel încât stările de spin ale celor doi atomi nu sunt perturbate — până când unul dintre noi nu efectuează o măsurătoare de spin, apăsând pe un buton. Atunci când unul dintre butoanele aflate pe vârfurile dodecaedrului este apăsat, acest lucru activează măsurătoarea de tip Stern-Gerlach asupra atomului din centrul său (să spunem că acest lucru se realizează prin folosirea unui câmp magnetic neomogen, ca în paragraful 5.10). Ne amintim că pentru spin 3/2, avem patru rezultate posibile corespunzând (în cazul în care aparatul este orientat să detecteze pe direcţia verticală în sus) celor patru stări ortogonale $|\uparrow\uparrow\uparrow>$ $|\downarrow\uparrow\uparrow>$ $|\uparrow\downarrow\downarrow>$ şi $|\downarrow\downarrow\downarrow>$ Aceste stări reprezintă cele patru locaţii posibile ale atomului după ce a întâlnit aparatul de măsură. Aranjamentul realizat de Tinichelele Chintesenţiale a fost ca, de fiecare dată când se apasă pe buton, aparatul de măsură să fie orientat în direcţia ce uneşte centrul cu acel buton. Soneria se va declanşa (răspuns DA) dacă atomul va fi în a doua din aceste patru locaţii posibile (figura 5.27). Asta revine la a spune (folosind notaţia pentru cazul 'sus') că starea $|\downarrow\uparrow\uparrow>$ determină răspunsul DA (soneria şi efectul pirotehnic calculat), pe când celelalte trei stări nu produc nici un răspuns. În cazul unui NU, cele trei locaţii care mai rămân sunt aduse împreună (de exemplu prin inversarea direcţiei cîmpului magnetic neomogen) fără ca distincţia dintre ele să aibă vreun efect exterior de perturbare; şi măsurătoarea este reluată, prin apăsarea unui alt buton şi selectarea unei alte direcţii. Observaţi că fiecare apăsare efectuează o măsurătoare *primitivă*, de tipul celor descrise în paragraful 5.13.

Pentru doi atomi de spin 3/2 proveniţi din starea de spin 0 şi cu starea $|\Omega>$, starea totală poate fi exprimată ca:

$$|\Omega> = |L\uparrow\uparrow\uparrow>|R\downarrow\downarrow\downarrow> - |L\uparrow\uparrow\downarrow>|R\downarrow\downarrow\uparrow> +$$
$$+ |L\uparrow\downarrow\downarrow>|R\downarrow\uparrow\uparrow> - |L\downarrow\downarrow\downarrow>|R\downarrow\downarrow\downarrow>$$

Dacă atomul meu este cel din partea dreaptă şi găsesc că starea lui este într-adevăr $|R\downarrow\uparrow\uparrow>$, deoarece soneria se declanşează la prima apăsare pe butonul cel mai de sus, atunci trebuie ca soneria de la dodecaedrul colegului meu să se declanşeze dacă el va apăsa prima dată butonul opus celui pe care am apăsat eu — iar starea atomului din dodecaedrul său este $|L\uparrow\downarrow\downarrow>$ Mai mult, dacă soneria mea nu sună pentru prima apăsare pe buton, atunci soneria lui nu va putea suna nici ea la

apăsarea pe butonul opus.

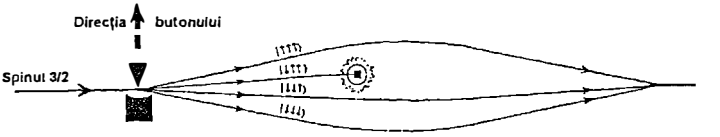


Fig. 5.27. Tinichelele Chintesenţiale au aranjat astfel ca, dacă un buton este apăsat, să se execute o măsurătoare de spin asupra atomului de spin 3/2 din interior, unde starea $|\downarrow\uparrow\uparrow>$ va activa soneria.

Acum mai trebuie să stabilim că proprietăţile (a) şi (b) (din paragraful 5.3), pe care 'Tinichelele Chintesenţiale' le garantează, se păstrează pentru aceast tip de măsurători primitive. În Anexa C sunt date anumite proprietăţi matematice ale descrierii Majorana a stărilor de spin, în particular pentru spinul 3/2. Vom simplifica discuţia dacă ne vom imagina sfera Riemann ca sfera ce trece prin toate vârfurile dodecaedrului — adică sfera circumscrisă dodecaedrului. Putem observa apoi că descrierea Majorana a stărilor DA pentru un buton apăsat în vârful P, este punctul P însuşi, luat împreună cu punctul P^* antipodul lui P — care este într-adevăr starea $|\downarrow\uparrow\uparrow>$ pentru P luat la polul nord. Putem nota această stare DA cu $|P^*PP>$

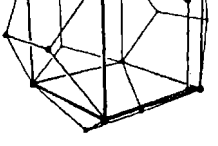


Fig. 5.28. Un cub poate fi plasat în interiorul dodecaedrului, folosind 8 din cele 20 de vârfuri. Vârfurile adiacente de pe cub se situează 'din doi în doi' pe dodecaedru.

O proprietate cheie a spinului 3/2 este că stările DA pentru măsurătorile primitive corespunzătoare apăsării pe butoanele unor vârfuri neadiacente ale dodecaedrului sunt ortogonale. De ce este aşa? Trebuie să stabilim că stările Majorana $|A^*AA>$ şi $|C^*CC>$ sunt de fapt ortogonale ori de câte ori A şi C sunt vârfuri neadiacente. Se va vedea din figura 5.28 că A şi C sunt neadiacente pe

dodecaedru ori de câte ori sunt vârfuri adiacente ale unui cub care se află în nteriorul dodecaedrului, cu acelaşi centru şi având opt vârfuri comune. Din Anexa C, paragraful final, $|A^*AA>$ şi $|C^*CC>$ sunt într-adevăr ortogonale ori de câte ori A şi C sunt vârfuri adiacente ale unui cub, astfel că rezultatul este stabilit.

Ce ne spun toate acestea? În primul rând, că apăsarea celor trei butoane adiacente celui notat SELECTAT, reprezintă trei măsurători *comutative* (paragraful 5.14), vârfurile în chestiune fiind toate neadiacente. Astfel, ordinea în care sunt apăsate nu are nici o importanţă pentru rezultatul final. Ordinea este de asemenea irelevantă pentru colegul meu din α-Centauri. Dacă el alege din întâmplare vârful opus celui ales de mine ca vârf SELECTAT, atunci cele trei apăsări posibile se exercită asupra unor vârfuri opuse celor alese de mine. Dar, aşa cum am mai spus, soneria se va declanşa în ambele cazuri la apăsarea butoanelor unor vârfuri opuse — indiferent de ordinea pe care o alegem — sau nu se va declanşa la nici una din primele trei apăsări. Toate acestea stabilesc proprietatea (a).

Ce putem spune despre (b)? Să observăm, mai întâi, că spaţiul Hilbert pentru starea de spin 3/2 este patru-dimensional, astfel încât cele trei posibilităţi ortogonale pentru care soneria se poate declanşa sunt, să spunem, $|A^*AA>$, $|C^*CC>$ şi $|G^*GG>$ — unde vârful SELECTAT pe dodecaedrul meu este B (vezi figura 5.29). Aceste posibilităţi nu reprezintă însă toate alternativele posibile. Ultima posibilitate apare atunci când soneria nu se declanşează la nici una din aceste apăsări şi avem o măsurătoare rezultantă nulă (soneria nu sună pentru nici una din cele trei încercări) care stabileşte că starea este (unica stare) ortogonală pe toate celelalte trei ($|A^*AA>$, $|C^*CC>$, $|G^*GG>$). Să notăm această stare cu *RST*, unde punctele *R*, *S*, *T* de pe sfera Riemann furnizează descrierea Majorana. Locaţia actuală a acestor trei puncte nu este uşor de stabilit (ele au fost localizate explicit de Jason Zimba (1993)). Pentru discuţia noastră nici nu contează foarte tare poziţia lor exactă. Tot ceea ce trebuie să ştim este că ele se află în poziţii determinate de geometria dodecaedrului în relaţie cu vârful SELECTAT *B*. Astfel, în particular, din considerente de simetrie, dacă aş fi ales vârful B^*, antipodal lui *B*, ca vârf SELECTAT în locul lui B, atunci starea $|R^*S^*T^*>$ — unde R^*, S^*, T^*, sunt vârfuri antipodale cu *R*, *S*, *T* — ar fi fost rezultatul măsurătorii nule (nedeclanşarea soneriei pentru nici unul din vârfurile A^*, C^*, G^* adiacente lui B^*).

Să presupunem acum că vârful selectat de colegul meu pe dodecaedrul lui este *B*, corespunzător exact vârfului selectat de mine. Dacă soneria nu se

declanşează la nici una din apăsările sale pe *A, C, G,* adiacente lui *B*, atunci măsurătorile sale (comutative) succesive obligă atomul *meu* să ajungă într-o stare ortogonală celor trei care corespund apăsărilor pe butoanele *opuse* pe dodecaedrul meu, A^*, C^*, G^*, adică, atomul meu este constrâns să fie în starea $|R^*S^*T^*>$ Totuşi, dacă şi soneria mea rămâne tăcută la toate cele trei apăsări ale butoanelor din *A, C* şi *G*, atunci starea va fi *RST*. Însă din proprietatea C1, Anexa C , rezultă că *RST* este ortogonală pe $R^*S^*T^*$, astfel că este imposibil ca ambele sonerii să nu se declanşeze la nici una din cele şase apăsări pe buton — ceea ce stabileşte proprietatea (b).

Astfel se explică cum au fost Tinichelele Chintesenţiale în stare să folosească nonlocalizarea cuantică pentru a garanta proprietăţile (a) şi (b). În paragraful 5.3 am observat că *dacă* cele două dodecaedre se comportă ca două obiecte independente, atunci proprietăţile 'de culoare' (c), (d) şi (e) rezultă imediat, ceea ce ne pune în faţa unei probleme insolubile (de 'colorare' a vârfurilor), cum arată Anexa B. Astfel, ceea ce au reuşit Tinichelele Chintesenţiale este ceva ce ar fi *imposibil* dacă cele două dodecaedre ar fi obiecte independente (odată expediate din fabrica firmei).

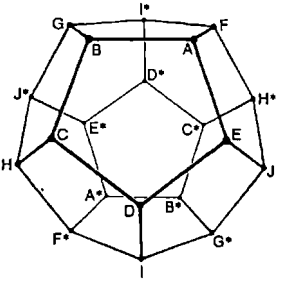


Fig. 5.29. Notaţia vârfurilor dodecaedrului pentru discuţiile din 5.18 şi Anexa B.

Nonlocalizarea cuantică nu este doar o curiozitate care ne spune că nu putem întotdeauna ignora efectele probabilistice ale mediului înconjurător asupra situaţiei fizice în discuţie. Atunci când efectele sale pot fi corespunzător izolate, este dimpotrivă vorba de ceva matematic precis, deseori cu o foarte bună organizare geometrică asociată.

Nici un fel de descriere în termenii unor obiecte independente nu poate

explica prezicerile formalismului mecanicii cuantice. În general nu putem avea o explicaţie de tipul 'şosetelor lui Bertlmann' pentru nonlocalizarea cuantică. Pe de altă parte, regulile standard ale evoluţiei cuantice U ne duc la concluzia că obiectele *trebuie* să rămână 'legate' în acet mod straniu, indiferent cât de îndepărtate sunt unul de altul. Doar cu **R** nonlocalizarea este anulată. Trebuie oare să credem că **R** este un proces 'real'? Dacă nu, nonlocalizarea ar trebui să persiste pentru totdeauna, chiar dacă este uneori ascunsă de complicaţiile excesive ale lumii actuale.

Înseamnă asta că orice lucru din univers este 'legat' de toate celelalte? Aşa cum am observat ceva mai devreme (în paragraful 5.17), nonlocalizarea cuantică este un efect care nu seamănă cu nimic din ceea ce ştim din fizica clasică, unde efectele tind să se atenueze cu creşterea distanţei, astfel încât nu avem nevoie să ştim ce se petrece în galaxia Andromeda pentru a explica lucrurile din laboratorul nostru de acasă. Nonlocalizarea cuantică ne spune că pare să existe un fel straniu de 'fantomatică acţiune la distanţă' care-i plăcea atât de puţin lui Einstein. Şi totuşi, este o 'acţiune' de un fel foarte subtil care nu poate fi folosită pentru transmiterea unor mesaje.

În ciuda faptului că nu reuşesc să ofere un potenţial pentru comunicarea la distanţă, efectele nonlocalizării cuantice nu pot fi ignorate. Atâta vreme cât nonlocalizarea persistă, nu putem considera, vorbind riguros, nici un obiect din Univers ca fiind complet independent. După părerea mea, această situaţie din fizica contemporană e departe de a fi satisfăcătoare. Nu există o explicaţie reală în teoria cuantică pentru faptul că, în practică, efectele nonlocalizătii *pot* fi ignorate. De ce nu este necesar să considerăm Universul ca o incredibil de complicată dezordine cuantică, fără nici o relaţie cu lumea clasică pe care o observăm? În practică, folosirea continuă a procedurii **R** este cea care 'rezolvă' nonlocalizarea, aşa cum se întâmplă cu atomii din centrul celor două dodecaedre atunci când sunt efectuate măsurătorile. Se pune deci întrebarea: este **R** un *proces fizic real*? Este nonlocalizarea cuantică într-adevăr eliminată? Sau totul trebuie explicat ca un fel de iluzie?

Voi încerca să discut aceste probleme în capitolul următor. După părerea mea, acestea sunt chestiunile centrale care trebuie abordate dacă vrem să găsim un rol noncomputabilităţii în fizică.

Anexa B

Imposibilitatea de a colora dodecaedrul

Amintiţi-vă problema pusă în paragraful 5.3: să arătăm că nu există nici un mod de a colora toate vârfurile dodecaedrului în ALB sau NEGRU, astfel încât două vârfuri neadiacente să nu poată fi amândouă ALBE, iar şase vârfuri adiacente unei perechi de vârfuri opuse să nu fie toate NEGRE. Simetria dodecaedrului ne va fi de un imens ajutor în cele ce urmează.

Să notăm vârfurile ca în figura 5.29. Aici, *A, B, C, D, E* sunt vârfurile unei feţe pentagonale, descrise ciclic, iar *F, G, H, I, J* sunt vârfurile adiacente lor, luate în aceeaşi ordine. Ca în paragraful 5.18, $A^*,\ldots J^*$ sunt vârfurile antipodale, luate în aceeaşi ordine. Să observăm mai întâi că, din cauza celei de-a doua proprietăţi, trebuie să fie cel puţin un vârf ALB undeva. Vom presupune că acest vârf alb este *A*.

Să presupunem, deocamdată, că vârful ALB A are, ca unul dintre vecinii cei mai apropiaţi, un alt vârf ALB — şi să considerăm că acesta este *B* (vezi figura 5.29). Cele zece vârfuri care le înconjoară pe acestea, *C, D, E, J,* H^*, *F,* I^*, *G,* J^* *H,* trebuie să fie toate NEGRE, deoarece ele sunt fiecare neadiacente cu *A* şi cu *B.* În continuare, examinăm cele şase vârfuri adiacente cu unul din vârfurile perechii antipodale *H,* H^* Între ele trebuie să fie unul alb, adică ori F^*, ori C^*, sau amândouă, trebuie să fie ALBE. Făcând acelaşi lucru pentru perechea antipodală *J,* J^*, concluzionăm că fie G^*, fie E^*, fie amândouă trebuie să fie albe. Ori acest lucru este imposibil, pentru că G^* şi E^* sunt neadiacente cu F^* şi C^*, ceea ce elimină şansa ca vârful *A* (ALB) să aibă un vecin imediat ALB — şi, din considerente de simetrie rezultă că nu este posibil să existe nici o pereche de vârfuri adiacente ALBE.

Astfel, vârful *A* ALB trebuie să fie înconjurat de vârfuri NEGRE, *B, C, D, E, J,* H^*, *F,* I^*, *G,* deoarece fiecare dintre acestea este fie adiacent lui *A* fie următor

adiacenței. Să examinăm acum cele șase vârfuri adiacente unuia din vârfurile perechii antipodale A, A^*. Vom trage concluzia că unul dintre vârfurile B^*, E^*, F^* trebuie să fie ALB, iar din simetria dodecaedrului rezultă că este indiferent care dintre ele este ALB, astfel încât putem considera că F^* este astfel. Observăm de asemenea că E^* și G^* sunt neadiacente cu F^* astfel încât trebuie să fie ambele NEGRE. De asemenea H trebuie să fie negru deoarece este adiacent cu F^* astfel încât ambele trebuie să fie negre, iar H trebuie să fie și el NEGRU pentru că este adiacent cu F^*, iar prin raționamentul anterior am eliminat posibilitatea unor vârfuri albe adiacente. Totuși, această colorare este imposibilă deoarece vârfurile antipodale J și J^* nu au acum decât vârfuri negre de jur împrejur, ceea ce închide discuția, arătând imposibilitatea *clasică* a dodecaedrelor magice.

Anexa C

Ortogonalitate între stările generale de spin

Descrierea Majorana a stărilor generale de spin nu este foarte familiară fizicienilor; totuși ea oferă o imagine utilă și geometric lămuritoare. Voi prezenta aici o variantă prescurtată a formulelor de bază și a implicațiilor lor geometrice; ca rezultat vom obține relațiile de ortogonalitate care se află la baza geometriei dodecaedrelor magice, așa cum ne era cerut în paragraful 5.18. Discuția mea va diferi considerabil de cea oferită de Majorana (1932), fiind mai apropiată de cele pe care le puteți găsi în Penrose (1994a) și Zimba și Penrose (1993).

Ideea este să considerăm mulțimea neordonată de n puncte de pe sfera Riemann ca reprezentând cele n rădăcini ale complexului polinomial de grad n și (esențial!) să folosim coeficienții acestui polinom ca și coordonate ale spațiului Hilbert $(n+1)$ dimensional al stărilor de spin pentru o particulă (masivă) de spin $\frac{1}{2}n$. Luând ca stări ale bazei diferitele rezultate posibile ale unei măsurători de spin în direcție verticală (ca în paragraful 5.10), le reprezentăm sub forma unor monoame (împreună cu un factor de normalizare corespunzător care ne asigură că

fiecare dintre aceste stări ale bazei este un vector unitate):

$|\uparrow\uparrow\uparrow \quad \uparrow\uparrow\rangle$ corespunde lui x^n

$|\downarrow\uparrow\uparrow\uparrow\ldots\uparrow\uparrow\rangle$ corespunde lui $n^{1/2}x^{n-1}$

$|\downarrow\downarrow\uparrow\uparrow\ldots\uparrow\uparrow\rangle$ corespunde lui $\{n(n-1)/2!\}^{1/2}x^{n-2}$

$|\downarrow\downarrow\downarrow\uparrow\ldots\uparrow\uparrow\rangle$ corespunde lui $\{n(n-1)(n-2)/3!\}^{1/2}x^{n-3}$

$|\downarrow\downarrow\downarrow\downarrow\ldots\downarrow\uparrow\rangle$ corespunde lui $n^{1/2}x$

$|\downarrow\downarrow\downarrow\downarrow\ldots\downarrow\downarrow\rangle$ corespunde lui 1

(Expresiile din paranteze sunt coeficienţi polinomiali)

Astfel, starea generală de spin $\frac{1}{2}n$:

$$z_0|\uparrow\uparrow\uparrow\ldots\uparrow\rangle + z_1|\downarrow\uparrow\uparrow \quad \uparrow\rangle + z_2|\downarrow\downarrow\uparrow \quad \uparrow\rangle + z_3|\downarrow\downarrow\downarrow\ldots\uparrow\rangle + \\ + z_n|\downarrow\downarrow\downarrow\ldots\downarrow\rangle$$

corespunde polinomului

$$p(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \quad + a_nx^n$$

unde

$$a_0 = z_0,$$

$$a_1 = n^{1/2}z_1,$$

$$a_2 = \{n(n-1)/2!\}^{1/2}z_2,$$

$$a_n = z_n.$$

Rădăcinile $x = \alpha_1, \alpha_2, \ldots\alpha_n$ ale lui $p(x) = 0$ furnizează n puncte pe sfera Riemann care definesc descrierea Majorana. Aceasta include posibilitatea punctului Majorana dat de $x = \infty$ (polul sud), care apare atunci când gradul polinomului $P(x)$ este mai mic decât n cu un număr dat de multiplicitatea acestui punct.

O rotaţie a sferei se obţine printr-o transformare conform căreia este făcută mai întâi înlocuirea:

$$x \rightarrow (\lambda x-\mu)(\bar{\mu}x+\bar{\lambda})^{-1}, \quad \text{unde} \quad \lambda\bar{\lambda}+\mu\bar{\mu} = 1$$

În al doilea pas, numitorii sunt eliminaţi prin înmulţirea întregii expresii cu $(\bar{\mu}x+\bar{\lambda})^n$. Astfel, polinoamele corespunzătoare rezultatelor măsurătorilor de spin (de tip Stern-Gerlach, de exemplu) într-o direcţie arbitrară pot fi obţinute, dându-se o expresie de forma:

$$c(\lambda x-\mu)^p(\bar{\mu}x+\bar{\lambda})^{n-p}$$

Punctele date de μ/λ şi $-\bar{\lambda}/\bar{\mu}$ sunt antipodale pe sfera Riemann şi

corespund direcţiei de măsurare a spinului, respectiv direcţiei opuse acesteia. (Ceea ce presupune o alegere corespunzătoare a fazelor pentru stările $|\uparrow\uparrow\uparrow...\uparrow>$, $|\downarrow\uparrow\uparrow...\downarrow>$, $|\downarrow\downarrow\uparrow\ \uparrow>$, $|\downarrow\downarrow\downarrow...\downarrow>$. Proprietăţile mai sus menţionate şi verificările lor detaliate sunt cel mai uşor de apreciat în cazul formalismului spinorial pentru spinor *2* (Penrose, Rindler, 1984, p. 162). Starea generală de spin *½n* este descrisă în termenii unui spinor *n*-valent, iar descrierea Majorana rezultă din descompunerea sa canonică ca produs simetric de vectori de spin.)

Punctul antipodal oricărui punct α de pe sferă este dat de $-1/\bar{\alpha}$ Astfel, dacă reflectăm toate punctele Majorana care sunt rădăcini ale ecuaţiei polinomiale

$$a(x) \equiv a_0 + a_1x + a_2x^2 + ... + a_{n-1}x^{n-1} + a_nx^n$$

în centrul sferei, obţinem rădăcinile polinomului

$$a^*(x) \equiv \bar{a}_n - \bar{a}_{n-1}x + \bar{a}_{n-2}x^2 - ... - (-1)^n\bar{a}_1x^{n-1} + (-1)^n\bar{a}_0x^n$$

Dacă avem două stări $|\alpha>$ şi $|\beta>$, date de polinoamele respective $a(x)$ şi $b(x)$, unde

$$b(x) \equiv b_0 + b_1x + b_2x^2 + ... + b_{n-1}x^{n-1} + b_nx^n$$

atunci produsul lor scalar va fi

$$<\beta|\alpha> = \bar{b}_0a_0 + \frac{1}{n}\bar{b}_1a_1 + \frac{2!}{n(n-1)}\bar{b}_2a_2 + \frac{3!}{n(n-1)(n-2)}\bar{b}_3a_3 + ... + \bar{b}_na_n$$

Această expresie este invariantă la rotaţiile pe sferă, aşa cum poate fi verificat folosind formulele de mai înainte.

Să aplicăm expresia produsului scalar cazului particular în care $b(x) = a^*(x)$, astfel încât avem de-a face cu două stări ale căror descrieri Majorana constau într-o pereche de puncte antipodale. Produsul lor scalar este (până la un semn)

$$a_0a_n - \frac{1}{n}a_1a_{n-1} + \frac{2!}{n(n-1)}a_2a_{n-2} - ... - (-1)^n\frac{1}{n}a_{n-1}a_1 + (-1)^na_na_0$$

Se poate vedea din acestea că dacă n este impar, atunci toţi termenii se anulează, astfel încât avem teorema următoare (starea cu descrierea Majorana *P, Q, ..., S* este notată cu $|PQ...S>$; Punctul antipodal lui *X* este notat prin X^*).

C.1 Dacă *n* este impar, starea $|PQR...T>$ este ortogonală pe $|P^*Q^*R^*...T^*>$.

Două alte proprietăţi rezultă din expresia generală a produsului scalar:

C.2 Starea $|PPP...P>$ este ortogonală pe fiecare dintre stările $|P^*AB...D>$.

C.3 Starea $|QPP...P>$ este ortogonală pe $|ABC...E>$ ori de câte ori

proiecţia stereografică, din P^* a punctului Q^* este centroidul proiecţiilor stereografice din P^* pe A, B, C, ..., D.

Centroidul unei mulţimi de puncte este centrul de gravitaţie al configuraţiei de puncte de masă egală situate în punctele respective. Proiecţia stereografică a fost descrisă în paragraful 5.10, figura 5.19. Pentru a demonstra C.3, rotim sfera până când P^* este în polul sud. Atunci starea $|QPP..P>$ este reprezentată de polinomul $x^{n-1}(x-\chi)$, unde χ defineşte punctul Q pe sfera Riemann. Formând produsul scalar cu starea reprezentată de polinomul $(x-\alpha_1)(x-\alpha_2)(x-\alpha_3)...(x-\alpha_n)$, unde descrierea Majorana este dată de α_1, α_2, α_3, ..., α_n, găsim că acesta se anulează atunci când

$$1 + n^{-1}\bar{\chi}(\alpha_1 + \alpha_2 + \alpha_3 +...+ \alpha_n) = 0$$

adică, atunci când $-1/\chi$ este egal cu $(\alpha_1 + \alpha_2 + \alpha_3 +...+ \alpha_n)/n$, tocmai centroidul punctelor date de

$$\alpha_1, \alpha_2, \alpha_3, ..., \alpha_n ,$$

în planul complex. Acesta este rezultatul stabilit de C.3.

Pentru a dovedi C.2, considerăm P ca fiind la polul sud. Atunci starea $|PPP...P>$ este reprezentată de constanta 1, considerată ca polinom de gradul n. Produsul scalar corespunzător se anulează acum atunci când

$$\alpha_1\alpha_2\alpha_3...\alpha_n = 0$$

adică atunci când cel puţin unul din α_1, α_2, α_3, ..., α_n se anulează punctul 0 din planul complex reprezentând polul nord P^*, ceea ce demonstrează C.2.

Rezultatul C.2 ne permite să interpretăm punctele Majorana în termeni fizici. Are ca implicaţii faptul că aceste puncte definesc direcţia în care o măsurătoare de spin de tip Stern-Gerlach ne dă probabilitatea zero de a găsi spinul în direcţia opusă celei de măsurare (vezi **MN**, p. 326, nota 10). Conţine de asemenea, ca un caz special, faptul că pentru spinul ½ (n=1), stările ortogonale sunt exact cele ale căror puncte Majorana sunt antipodale.

Rezultatul C.3 ne permite să deducem interpretarea geometrică generală ortogonalităţii în cazul spinului 1 (n=2). Un caz particular interesant apare când cele două stări sunt reprezentate ca două perechi de puncte antipodale ale căror legătură este dată de liniile perpendiculare care trec prin centrul sferei. În cazul spinului 3/2 (n=3), C.3 furnizează (cu C.1) tot ce avem nevoie pentru paragraful 5.18 (o interpretare geometrică a ortogonalităţii în cazul general va fi dată în altă parte). Cazul particular al lui C.3, de care avem nevoie pentru paragraful 5.18,

apare atunci când P şi Q sunt vârfuri adiacente ale unui cub înscris în sfera Riemann, astfel încât PQ şi Q^*P^* sunt muchii opuse. Lungimile PQ^* şi QP^* sunt de $\sqrt{2}$ ori cele ale PQ şi P^*Q^*. Rezultă din C.3, dintr-o geometrie simplă, că stările $|P^*PP>$ şi $|Q^*QQ>$ sunt ortogonale.

Note şi referinţe bibliografice

1. Penrose (1993b, 1994a), Zimba şi Penrose (1993)
2. Sugestia iniţială pentru un experiment bine definit provine de la Clauser şi Horne, (1974) şi Clauser, Horne şi Shimony, (1978).
3. Primele experimente care au indicat o confirmare pozitivă a aşteptărilor cuantice nonlocale au fost cele ale lui Freedman şi Clauser, (1972), urmate câţiva ani mai târziu de rezultatele mult mai importante ale lui Aspect, Grangier şi Roger (1982). (vezi şi Aspect şi Grangier, (1986)).
4. Există un alt tip de explicaţie 'clasică' posibilă pentru efectele EPR particulare care au fost observate până acum de Aspect şi alţii. Această explicaţie sugerată — colaps retardat este datorată lui Euan Squires (1992a) şi ia în considerare faptul că poate exista o întârziere semnificativă în efectuarea propriu-zisă a măsurătorii dacă detectorii sunt în locaţii separate. Această sugestie trebuie considerată în cadrul unei teorii oarecare în mod necesar una neconvenţională, la fel ca cele pe care le veţi întâlni în paragrafele 6.9 şi 6.12 care să producă nişte predicţii bine definite, de pildă asupra timpului probabil la care cele două măsurători vor avea loc în mod obiectiv. Datorită influenţelor aleatoare care se exercită asupra acestor valori de timp, s-ar putea întâmpla ca unul dintre detectori să efectueze măsurătoarea semnificativ mai devreme decât celălalt atât de devreme încât s-ar putea întâmpla ca un semnal luminos să treacă de la primul detector la al doilea şi să-l 'informeze' pe acesta care a fost rezultatul măsurătorii primului detector.

   Conform acestui punct de vedere, ori de câte ori are loc o măsurătoare cuantică, ea este acompaniată de o *undă de informaţie* care se deplasează cu viteza luminii. Acest gen de lucruri este conform cu teoria relativităţii clasice (vezi paragraful 4.4), dar e în dezacord cu predicţiile teoriei cuantice ,pentru evenimente care au loc la distanţe mari. În particular, dodecaedrele magice din paragraful 5.3 nu pot fi explicate în termenii colapsului retardat. Desigur, nici un astfel de experiment nu a fost încă efectuat şi unii dintre ar putea adopta părerea că el ar contrazice predicţiile teoriei cuantice. O obiecţie ceva mai serioasă este cea

conform căreia colapsul retardat va întâlni numeroase dificultăţi în cazul altor tipuri de măsurătoare cuantică şi va conduce la o violare a legilor de conservare standard. De exemplu, atunci când un atom radioactiv emite o particulă încărcată — să spunem o particulă alfa — va fi posibil pentru doi detectori separaţi aflaţi la o distanţă considerabilă să înregistreze aceeaşi particulă alfa, violând simultan legile de conservare pentru energie, sarcină electrică şi număr barionic! (Pentru o separare destul de mare, *unda de informaţie* care vine de la primul detector nu va avea destul timp pentru a avertiza al doilea detector că nu poate observa aceeaşi particulă alfa!). Totuşi, aceste legi de conservare vor fi respectate 'în medie' şi nu ştiu să existe nici o observaţie actuală care să le contrazică. Pentru o tratare recentă a statutului colapsului retardat, vezi Horne, (1994).

5. Am fost informat de Abner Shimony că Kochen şi Specker erau deja la curent cu formularea EPR posibilă pentru experimentul lor.
6. Pentru alte exemple, cu configuraţii geometrice diferite, vezi Peres (1990, 1991), Mermin (1990), Penrose (1994a).
7. Cea mai eficientă oglindă semiargintată nu este, de fapt, argintată deloc, ci este o bucată subţire de material transparent, de grosimea potrivită lungimii de undă a luminii. Îşi va atinge efectul printr-o combinaţie complicată de reflexii interne şi transmisii, astfel încât razele finale reflectată şi transmisă sunt egale în intensitate. Rezultă din natura 'unitară' a transformărilor între razele transmisă şi reflectate emergente că trebuie să aibă loc o rotaţie de fază de un sfert de lungime de undă, dat fiind factorul i cerut. Vezi Klein şi Furtak (1986) pentru o discuţie ceva mai complicată.
8. Vezi de exemplu Dirac (1947), Davies (1984).
9. Există un anumit grad de arbitrar în alegerea unui factor de fază pe care am adoptat-o aici pentru starea reflectată. Aceasta depinde în parte de ce tip de oglindă este folosit. De fapt, spre deosebire de oglinda 'semiargintată' la care ne-am referit în nota 7 (şi care nu era de fapt argintată deloc) putem considera aceste două oglinzi ca fiind complet argintate. Factorul i pe care l-am adoptat aici este un fel de compromis, realizând un acord superficial cu factorul obţinut în cazul unei reflexii pe oglinzile semiargintate. De fapt, nu contează cu adevărat ce factor este adoptat pentru reflexia pe oglinzile complet argintate, atâta timp cât suntem consistenţi în ceea ce facem, pentru ambele oglinzi despre care este vorba.
10. De exemplu, Kochen şi Specker (1967) şi referinţele date în nota 6.

6. TEORIA CUANTICĂ ŞI REALITATEA

6.1. Este R un proces real?

În capitolul precedent am încercat să ne familiarizăm cu enigmaticele mistere **Z** ale mecanicii cuantice. Cu toate că nu toate aceste fenomene au fost testate experimental — ca, de exemplu, non-separabilitatea cuantică la distanţe de câţiva ani lumină [1] — există deja destul sprijin experimental pentru a susţine că misterele **Z** reprezintă chestiuni care trebuie luate în serios şi că ele sunt aspecte ale comportamentului conştituenţilor lumii în care trăim.

Acţiunile lumii fizice la nivel cuantic sunt, într-adevăr, profund contraintuitive şi în multe feluri oarecum diferite de comportamentul 'clasic' care pare să opereze la niveluri mai familiare ale exeprienţei noastre. Comportarea cuantică a Universului include efecte de non-separabilitate la distanţă de mai mulţi metri, cel puţin atâta vreme cât în discuţie se află obiecte cuantice ca electroni, fotoni, atomi şi molecule. Contrastul dintre straniul comportament *cuantic* al lucrurilor 'mici', chiar separate la distanţe mari şi comportamentul *clasic* ceva mai familiar al lucrurilor ceva mai mari se află la baza problemei misterelor **X** ale teoriei cuantice. S-ar putea oare întâmpla să existe două tipuri de legi fizice, unele care operează la nivelul fenomenelor iar altele care operează la nivel microscopic?

O astfel de idee este oarecum în contradicţie cu lucrurile la care ne-am aştepta din partea fizicii. Într-adevăr, una dintre cele mai profunde descoperiri ale dinamicii secolului XVII, datorate lui Galilei şi lui Newton, a fost ideea că

mişcările corpurilor cereşti ascultă de aceleaşi legi fizice ca cele care operează pe pământ.

Încă din timpul Greciei antice, dacă nu dinainte, se credea că există un set de legi pentru ceruri şi un alt set de legi, complet diferite, pentru ceea ce se întâmplă pe pământ. Galilei şi Newton ne-au învăţat că legile sunt aceleaşi — o intuiţie fundamentală, esenţială pentru progresul ştiinţei. Şi totuşi (aşa cum a subliniat profesorul Ian Percival de la Universitatea din Londra), cu teoria cuantică lucrurile par să se fi întors la o schemă ca cea a Greciei antice, cu un corp de legi care operează la nivel clasic şi un altul, foarte diferit, la nivel cuantic.

După părerea mea — şi nu este numai părerea mea, ci una împărtăşită de o minoritate vizibilă de fizicieni — acest stadiu al fizicii nu poate fi altceva decât un moment de criză. Am putea anticipa faptul că găsirea unor legi cuantice/clasice corespunzătoare care să opereze în mod uniform la *orice* nivel va anunţa un progres ştiinţific de dimensiuni comparabile celui iniţiat de Galilei şi Newton.

Cititorul s-ar putea întreba pe bună dreptate dacă într-adevăr teoria cuantică standard ne oferă o imagine care nu este în stare să explice fenomenele clasice. Mulţi mă vor contrazice atunci când afirm că aceasta este situaţia, spunând că sistemele fizice mari sau complicate şi care acţionează în întregime după legile cuantice, se vor comporta exact ca şi sistemele clasice (ca o foarte bună aproximaţie a sistemelor clasice). Vom încerca mai întâi să vedem dacă această aserţiune este credibilă, dacă, într-adevăr, comportamentul aparent 'clasic' al obiectelor macroscopice rezultă din comportamentul cuantic al constituenţilor macroscopici. Iar dacă vom găsi că nu este, vom încerca să vedem încotro trebuie să ne îndreptăm pentru a ajunge la un punct de vedere coerent care să aibă sens pentru *toate* nivelele realităţii.

Trebuie să vă avertizez, totuşi, că întreaga chestiune este subiect de controversă. Există mai multe puncte de vedere diferite şi ar fi prostesc să încerc să dau un rezumat comprehensiv pentru toate, ca să nu mai vorbim de detalierea celor pe care le cred neplauzibile sau de-a dreptul false. Cer deci indulgenţa cititorului pentru faptul că punctele de vedere pe care le prezint vor fi descrise din prespectiva mea. Este deci inevitabil să fiu nedrept cu punctele de vedere cu care nu sunt de acord şi îmi cer scuze în avans pentru posibilele nedreptăţi pe care urmează să le comit.

Când încercăm să găsim dimensiunile clare ale nivelului la care acţiunea

cuantică — caracterizată de persistenţa superpoziţiilor diferitelor alternative — cedează locul acţiunii *clasice* (în care nu avem superpoziţii) există o dificultate fundamentală. Ea rezultă din faptul că procedura **R** este o procedură 'alunecoasă' din punct de vedere observaţional şi care ne împiedică să identificăm un nivel/dimensiune la care ea are loc — şi acesta este unul dintre motivele pentru care mulţi fizicieni nu privesc **R** ca pe un fenomen real. La nivel experimental, nu pare să existe vreo diferenţă dacă alegem să aplicăm **R** sau nu, atâta timp cât ne aflăm la un nivel mai mare decât cel la care au fost observate efecte de interferenţă cuantică, dar nu mai mare decât cel la care putem percepe direct că alternativele clasice iau într-adevăr locul superpoziţiilor clasice (cu toate că unii susţin că superpoziţiile persistă chiar la acest nivel).

Cum am putea decide la ce nivel are lor **R** în mod *actual?* (presupunând că are loc în mod fizic). Este greu de văzut cum am putea răspunde unei astfel de întrebări printr-un experiment fizic. Dacă **R** este un proces fizic real, există un mare număr de nivele posibile la care poate avea loc, de la nivelele microscopice la care interferenţele cuantice au fost deja observate, la cele mult mai mari la care este percepută în mod actual comportarea clasică. Mai mult, aceste diferenţe de 'nivel' nu par să se refere la dimensiune, deoarece am văzut mai sus (în paragraful 5.4) că efectele non-separabilităţii cuantice se pot întinde la distanţe de mulţi metri. Vom vedea mai târziu că *diferenţa de energie* furnizează o măsură mai bună a nivelului cuantic decât dimensiunile fizice. Oricum ar sta lucrurile, însă, locul în care 'se aleg lucrurile' la capătul macroscopic al realităţii este furnizat de percepţiile noastre. Este o chestiune bizară din punctul de vedere al teoriei fizice, deoarece nu ştim cu adevărat care sunt, în creier, procesele fizice asociate percepţiei. Totuşi, natura fizică a acestor procese pare să ne furnizeze limita de sus pentru orice teorie a unui proces **R** *real*. Această condiţie permite încă existenţa unui mare interval între cele două extreme şi suficient loc pentru numeroase atitudini diferite privind ce anume se întâmplă *cu adevărat* atunci când **R** are loc.

Una dintre problemele principale priveşte 'realitatea' formalismului cuantic — sau chiar a lumii la nivel cuantic. În legătură cu acest lucru nu pot să nu citez o replică pe care mi-a dat-o odată, la o petrecere, profesorul Bob Wald de la Universitatea din Chicago:

Dacă crezi cu adevărat în mecanica cuantică, e imposibil s-o iei în serios.

Îmi pare că această replică exprimă ceva profund privind teoria cuantică şi

atitudinea oamenilor în faţa ei. Cei mai vehemenţi susţinători ai faptului că teoria cuantică nu are nevoie de modificare tind să creadă, în acelaşi timp, că ea *nu* reprezintă comportarea actuală a unei lumi 'reale'

Niels Bohr, unul dintre pionierii dezvoltării şi interpretării teoriei cuantice a fost unul dintre cei mai extremişti susţinători ai punctului de vedere de mai sus. El pare să fi considerat că vectorul de stare nu este mai mult decât o convenţie utilă la calculul probabilităţilor pentru rezultatele 'măsurătorilor' care pot fi efectuate asupra unui sistem. Vectorul de stare însuşi *nu* era considerat ca furnizând o descriere obiectivă de orice fel a *realităţii* la nivel cuantic, ci ca reprezentând doar 'cunoaşterea noastră' asupra sistemului. Bohr considera îndoielnică şi lipsită de sens chiar şi aplicarea conceptului de 'realitate' la nivel cuantic. Iar Bohr era cineva care "credea cu adevărat în mecanica cuantică" — pe când descrierea dată de el vectorului de stare pare să indice că *nu* trebuie 'luat în serios' ca descriere a realităţii fizice la nivel cuantic.

Alternativa la acest punct de vedere cuantic este, în linii mari, să considerăm că vectorul de stare furnizează o descriere matematică a lumii *reale* la nivel cuantic — o lume care evoluează cu un extraordinar grad de precizie, deşi, poate, nu cu o totală acurateţe, conform regulilor matematice date de ecuaţiile teoriei. Mi se pare că avem în acest domeniu larg două drumuri principale. În primul rând, sunt cei care consideră că procedura **U** descrie întreaga evoluţie a stării cuantice. Rezultă că procedura **R** este considerată un fel de iluzie, convenţie, aproximaţie şi *nu* trebuie considerată ca făcând parte din evoluţia *actuală* a realităţii descrise de starea cuantică. Aceste persoane vor fi duse în direcţia a ceea ce se numeşte *interpretarea* (interpretările) *lumilor multiple (many-worlds)*, sau *interpretarea Everett* [2]. Voi explicita acest punct de vedere în continuare. Pe de altă parte, cei care 'iau în serios' formalismul cuantic sunt cei care cred că atât **U** cât şi **R** reprezintă comportarea fizică *actuală* a unei lumi fizice reale descrise cu ajutorul vectorului de stare. Pentru aceştia, însă, devine greu de crezut că teoria cuantică este completă şi precisă la toate nivelele. Pentru că acţiunea lui **R** este în contradicţie cu multe dintre proprietăţile lui **U**, în particular cu *linearitatea* sa. În acest sens, ar fi vorba despre cei care "nu cred cu adevărat în mecanica cuantică" În secţiunea care urmează voi discuta pe larg aceste chestiuni.

6.2. Puncte de vedere de tip *many worlds*

Să vedem mai întâi cât de departe putem ajunge dacă urmărim celălalt drum 'realist' care duce în ultimă instanţă la o intepretare numită interpretarea lumilor multiple (many-worlds). În această interpretare, vectorul de stare, evoluând sub acţiunea lui **U**, este considerat ca având realitate fizică. Acest lucru ne obligă să acceptăm că obiectele clasice, ca, de exemplu, mingile de golf sau oamenii trebuie de asemenea să se comporte conform legilor superpoziţiei liniare cuantice. S-ar putea sugera că acest lucru nu trebuie neapărat să provoace dificultăţi câtă vreme aceste stări suprapuse par să fie foarte rar întâlnite la nivel clasic. Totuşi, există o mare problemă şi aceasta este *liniaritatea* lui **U**. Sub acţiunea lui **U**, factorii de pondere pentru stările suprapuse rămân aceiaşi indiferent care este cantitatea de materie care intră în discuţie. Procedura **U** nu permite, prin ea însăşi, superpoziţiilor să devină 'ne-superpoziţii' doar pentru că un sistem devine mare sau complicat. Astfel de superpoziţii nu vor tinde deci să 'dispară' pentru obiectele clasice, astfel încât ne aflăm în faşa unei situaţii în care obiectele clasice ar trebui, în fapt, să apară în stări suprapuse manifeste. Problema care trebuie rezolvată este: de ce superpoziţiile de alternative la nivel macroscopic nu afectează percepţiile noastre asupra lumii la nivel clasic?

Să încercăm să înţelegem cum ar răspunde susţinătorii interpretării lumilor multiple. Să considerăm o situaţie cum este cea discutată în paragraful 5.17, în care un detector de fotoni, descris cu ajutorul stării $|\Psi>$, întâlneşte un foton aflat în superpoziţia de stări $|\alpha>+|\beta>$, unde $|\alpha>$ afectează detectorul, iar $|\beta>$ îl lasă neperturbat (poate că un foton, emis de o sursă oarecare a interacţionat cu o oglindă semiargintată, iar $|\alpha>$ şi $|\beta>$ reprezintă părţile transmisă şi reflectată ale stării fotonului.) Nu trebuie să chestionăm aplicabilitatea conceptului de vector de stare la un obiect clasic de tipul detectorului pentru că în intepretarea lumilor multiple vectorii de stare sunt consideraţi reprezentări adecvate ale realităţii la toate nivelele. Astfel, $|\Psi>$ poate descrie întregul detector şi nu doar anumite părţi ale lui la nivel cuantic ca în paragraful 5.17. Amintiţi-vă că, în 5.17, după ce întâlnea fotonul, starea detectorului şi a fotonului evoluau de la produsul $|\Psi>(|\alpha>+|\beta>)$ la starea neseparabilă

$$|\Psi_Y> + |\Psi_N>|\beta'>$$

Această stare *totală* neseparabilă este considerată ca reprezentând *realitatea*

situaţiei. Nu vom spune că detectorul a primit şi absorbit fotonul (starea $|\Psi_Y>$) *sau* că detectorul nu a înregistrat fotonul care rămâne liber (starea $|\Psi_N>|\beta'>$), ci vom susţine că *ambele* alternative coexistă în superpoziţie, ca parte a unei realităţi totale în care toate superpoziţiile sunt păstrate. Putem duce mai departe această descriere şi ne putem imagina că un observator uman examinează detectorul pentru a vedea dacă a înregistrat sau nu fotonul. Experimentatorul uman, înainte de a examina detectorul, trebuie să se afle de asemenea într-o stare cuantică, să spunem $|\Sigma>$, astfel încât avem o stare produs combinată în acest stadiu şi anume

$$|\Sigma>(|\Psi_N> + |\Psi_N>|\beta'>)$$

După examinarea stării detectorului, observatorul fie percepe detectorul ca înregistrând particula şi absorbind fotonul (starea $|\Sigma_Y>$), fie îl percepe ca neînregistrându-l (starea ortogonală $|\Sigma_N>$). Dacă presupunem că observatorul nu interacţionează cu detectorul după ce-l observă, vom avea următoarea formă pentru starea vectoare care descrie situaţia:

$$|\Sigma_Y>|\Psi'_Y> + |\Sigma_N>|\Psi'_N>|\beta''>$$

Există acum două stări diferite (ortogonale) ale observatorului, fiecare implicate în starea totală a sistemului. Conform uneia dintre acestea, observatorul percepe detectorul înregistrând fotonul, iar aceasta este acompaniată de starea detectorului în care fotonul a fost cu adevărat înregistrat. Conform celeilalte, observatorul este în starea în care a perceput că detectorul nu înregistrează fotonul, iar aceasta este acompaniată de starea detectorului care nu a interacţionat cu fotonul şi cu starea fotonului mişcându-se liber. În interpretarea lumilor posibile, atunci, vor exista două instanţe (copii) ale 'sinelui' observatorului, coexistente cu starea totală şi având, fiecare, percepţii diferite despre lumea dimprejur. Starea actuală a lumii care acompaniază fiecare dintre copiile observatorului va fi consistentă cu percepţiile acelei copii.

Putem generaliza aceste rezultate la situaţia mai 'realistă' în care ar exista un uriaş număr de alternative cuantice diferite apărând în mod continuu de-a lungul istoriei universului în locul celor două din exemplul de mai sus. Astfel, conform punctului de vedere al lumilor multiple, starea totală a universului va cuprinde într-adevăr multe lumi 'diferite' şi multe instanţe diferite ale fiecărui observator uman. Fiecare astfel de copie va percepe o lume care este consistentă cu propriile sale percepţii şi se argumentează în continuare că acestea sunt toate lucrurile de care avem nevoie pentru o teorie satisfăcătoare. Procedura **R** nu este în această

interpretare decât o *iluzie* care apare ca o consecinţă a felului în care un observator macroscopic precepe într-o lume cuantică non-separabilă.

În ce mă priveşte, trebuie să spun că găsesc acest punct de vedere extrem de nesatisfăcător. Şi nu este vorba atât de marea lipsă de economie pe care această imagine o susţine — deşi şi aceasta este o trăsătură cel puţin îngrijorătoare. Obiecţia mai serioasă este că interpretarea lumilor posibile nu furnizează cu adevărat o soluţie a 'problemei măsurătorii' pe care trebuia s-o soluţioneze.

Această *problemă a măsurătorii cuantice* înseamnă de fapt a înţelege cum poate apărea procedura **R** ca o proprietate a comportamentului la scară mare a sistemelor cuantice care evoluează conform lui **U**. Problema nu se rezolvă indicând pur şi simplu o cale posibilă în care un comportament de tip **R** poate fi pus de acord cu restul teoriei. Ar trebui să avem o teorie care să furnizeze o anumită înţelegere a *circumstanţelor* în care apare (iluzia ?) **R**. Mai mult, ar trebui să avem o explicaţie a remarcabilei *precizii* implicată în **R**. În general, atunci când este vorba despre precizia teoriei cuantice oamenii se gândesc la ecuaţiile dinamice, adică al **U**. Dar **R** însuşi este foarte precis în privinţa predicţiei probabilităţilor şi câtă vreme nu suntem în stare să înţelegem cum funcţionează aceasta, nu vom avea o teorie satisfăcătoare.

În lipsa altor ingrediente, interpretarea lumilor multiple nu rezolvă cu adevărat nici una dintre aceste probleme. Fără o teorie despre cum anume poate o 'fiinţă perceptivă' să dividă lumea în alternative ortogonale, nu avem nici un motiv să ne aşteptăm ca o astfel de fiinţă să nu perceapă superpoziţii liniare de mingi de golf sau elefanţi în poziţii total diferite (să menţionăm că simpla *ortogonalitate* a 'stărilor perceptibile', precum $|\Sigma_Y\rangle$ şi $|\Sigma_N\rangle$ nu poate determina singular aceste stări. Comparaţia care se poate face este cu cazul lui $|L\leftarrow\rangle$ şi $|L\rightarrow\rangle$ ca opuse lui $|L\uparrow\rangle$ şi $|L\downarrow\rangle$ din discuţia EPR din paragraful 5.17. În fiecare caz, perechea de stări este ortogonală, aşa cum sunt $|\Sigma_Y\rangle$ şi $|\Sigma_N\rangle$, dar nu există nimic care să ne permită să alegem o pereche în defavoarea alteia). Mai mult, interpretarea lumilor multiple nu oferă nici o explicaţie pentru minunat de exacta regulă conform căreia modulul pătrat al factorilor de pondere complecşi devin, în mod miraculos, probabilităţi relative [3] (vezi pentru comparaţie şi discuţiile din paragrafele 6.6 şi 6.7.).

6.3. Neluându-l pe $|\psi>$ în serios

Există mai multe versiuni ale punctului de vedere conform căruia vectorul de stare $|\psi>$ nu este privit ca furnizând o imagine actuală a realităţii fizice la nivel cuantic. În loc de asta, $|\psi>$ este considerat un instrument de calcul, util pentru calcularea probabilităţilor, sau ca o expresie a 'stării de cunoaştere' a observatorului privind sistemul fizic. Uneori, $|\psi>$ este considerat ca reprezentând nu starea unui sistem individual, ci un *ansamblu* de sisteme fizice posibile similare. Se argumentează adesea că un vector de stare complicată non-localizată $|\psi>$ se va comporta 'în toate scopurile practice' (for all practical purposes, FAPP după expresia lui John Bell, [4]) în acelaşi fel ca un astfel de ansamblu de sisteme fizice şi că acestea sunt toate informaţiile de care fizicianul are nevoie privind problema măsurătorii. Uneori se argumentează chiar că $|\psi>$ *nu poate* descrie realitatea la nivel cuantic pentru că nu are sens să vorbim despre o 'realitate' a lumii la acel nivel. În acest caz, realitatea este înţeleasă ca fiind alcătuită exclusiv din rezultatele măsurătorilor.

Pentru cei ca mine (şi pentru Einstein şi Schrödinger de asemenea sunt, deci, într-o companie onorabilă), nu are sens folosirea termenului de realitate doar pentru obiectele pe care le putem percepe, ca, de exemplu, (anumite tipuri de) aparate de măsură. Fără îndoială că la nivel cuantic lumea este stranie şi nefamiliară; nu este însă 'nereală' Cum ar putea fi obiectele reale alcătuite din constituenţi nereali? Mai mult, legile matematice care guvernează lumea cuantică sunt remarcabil de precise tot atât de precise ca şi ecuaţiile mai familiare care controlează comportamentul obiectelor macroscopice în ciuda imaginilor vagi pe care le creează descrieri de tipul 'fluctuaţiilor cuantice' sau 'principiul de nedeterminare'.

Şi totuşi, chiar dacă acceptăm că la nivel cuantic trebuie să existe o anumită realitate, putem încă avea îndoieli privind felul în care aceasta ar putea fi descrisă de un vector de stare. Există numeroase şi variate obiecţii împotriva 'realităţii' vectorului de stare $|\psi>$. În primul rând, $|\psi>$ pare să treacă din timp în timp prin acele misterioase 'salturi' non-locale şi discontinue pe care le-am notat prin litera **R**. Nu acesta este felul în care trebuie să se comporte o descriere fizică acceptabilă a lumii, în special când avem o minunată ecuaţie continuă, **U**, pentru a controla felul în care $|\psi>$ evoluează (în majoritatea timpului). Şi totuşi,

aşa cum am văzut, U însuşi ne conduce la dificultăţi şi enigme de tipul lumilor posibile. Deci, dacă vrem o imagine care să semene cât de cât cu universul actual pe care îl vedem în jurul nostru, avem nevoie de ceva de natura lui **R**.

O altă obiecţie la realitatea lui $|\psi>$ este că tipul de alternaţie U, **R**, U, **R**, U, **R**,... folosită în teoria cuantică nu este o descriere simetrică în timp (deoarece **R** este cel care determină punctul de pornire al fiecărei acţiuni **U**, nu însă şi punctul final) şi că există o altă descriere complet echivalentă în care evoluţiile temporale ale lui **U** sunt inversate (vezi **MN**, figurile 8.1, 8.2). De ce ar trebui să considerăm mai degrabă una dintre acestea ca reprezentând realitatea şi nu cealaltă? Există chiar puncte de vedere conform cărora *ambii* vectori de stare, unul evoluând înainte în timp, altul înapoi, trebuie consideraţi ca părţi coexistente ale descrierii realităţii fizice (Costa de Beauregars, 1989, Werbos, 1989, Aharonov şi Vaidman, 1990). Cred că este posibil să existe ceva profund semnificativ la baza acestor consideraţii, însă pentru moment nu vreau să intru în aceste chestiuni. Voi reveni la ele şi la alte probleme înrudite, în paragraful 7.12.

Una dintre cele mai frecvente obiecţii împotriva realităţii lui $|\psi>$ este că nu e direct 'măsurabil' – în sensul în care, dacă cuiva i se arată o stare complet necunoscută, nu există nici o modalitate experimentală de a determina care este vectorul de stare asociat. Să luăm cazul unui atom de spin 1/2, de exemplu. Amintiţi-vă (paragraful 5.10, figura 5.19) că fiecare stare de spin posibilă este caracterizată de o direcţie particulară în spaţiu. Dacă însă nu avem idee care este acea direcţie, nu avem nici un mod de a o determina. Tot ce putem face este să stabilim aparatul de măsură într-o direcţie oarecare şi să punem întrebarea: este spinul orientat în această direcţie (DA) sau în direcţia opusă (NU)? Oricare este starea de spin iniţială, direcţia sa în spaţiul Hilbert este proiectată fie pe spaţiul DA, fie pe spaţiul NU, cu anumite probabilităţi. Şi în acest punct am pierdut deja cea mai mare parte a informaţiei privind care anume a fost starea iniţială 'actuală' a spinului. Tot ce putem obţine, dintr-o măsurătoare a direcţiei spinului, pentru un atom de spin ½ este *un bit* de informaţie (adică răspunsul la o întrebare da/nu), pe când stările posibile de spin formează un continuum, care are nevoie de un număr infinit de biţi de informaţie pentru a fi determinat în mod precis.

Toate acestea sunt adevărate, însă considerarea lui $|\psi>$ ca ceva 'nereal' (care înglobează cunoştinţele noastre asupra sistemului fizic considerat) nu este mai uşoară. Dimpotrivă, o consider foarte greu de acceptat, în special pentru că pare

să existe ceva foarte subiectiv privind acest rol al 'cunoaşterii' A cui cunoaştere, este până la urmă, cea la care se face referinţă? Sigur nu a mea. Am foarte puţină cunoaştere actuală a stărilor vectoare relevante pentru comportarea detaliată a obiectelor care mă înconjoară. Şi toate acestea îşi continuă foarte bine acţiunile lor organizate, în totală necunoştinţă de cauză relativ la ce anume este 'cunoscut' despre starea vectoare, sau cine anume este cel care o cunoaşte. Folosesc experimentatori diferiţi, cu o diferită cunoaştere despre sistemul fizic, stări vectoare diferite pentru a descrie sistemul? Nu într-un mod semnificativ; ei ar putea-o face, eventual, dacă aceste diferenţe s-ar referi la trăsături ale experimentului care sunt neesenţiale pentru rezultat.

Unul dintre cele mai puternice motive pentru a respinge un astfel de punct de vedere subiectiv privind realitatea lui [5] $|\psi\rangle$ vine din faptul că oricare ar fi $|\psi\rangle$, există întotdeauna, în principiu, cel puţin, o *măsurătoare primitivă* (conform paragrafului 5.13) al cărui spaţiu DA este alcătuit din vectorul în spaţiul Hilbert determinat de $|\psi\rangle$. Ideea este că starea fizică $|\psi\rangle$ (determinată de multiplii complecşi ai lui $|\psi\rangle$) este *unic* determinată de faptul că rezultatul DA, pentru această stare, este *sigur*. Nici o altă stare fizică nu are această proprietate. Pentru orice altă stare, ar exista doar o anumită probabilitate ca rezultatul să fie DA, însă nu certitudine, iar NU poate de asemenea apare. Astfel, cu toate că nu există măsurătoare care să ne spună ce este $|\psi\rangle$ cu adevărat, starea fizică $|\psi\rangle$ este unic determinată de rezultatul unei măsurători care ar putea fi efectuate asupra ei. Din nou este o problemă de contrafactuale (paragrafele 5.2, 5.3), dar am văzut cât sunt de importante contrafactualele pentru aşteptările teoriei cuantice.

Pentru a spune lucrurile mai caţegoric, să ne imaginăm că avem un sistem cuantic într-o stare cunoscută, să spunem $|\phi\rangle$ Putem calcula că, după timpul t, starea va evolua, sub acţiunea lui **U**, într-o altă stare $|\psi\rangle$ De exemplu, $|\phi\rangle$ ar putea reprezenta starea de 'spin sus' ($|\phi\rangle = |\uparrow\rangle$) a unui atom de spin ½ şi putem presupune ca a fost pus în această stare de către acţiunea unei măsurători precedente. Să presupunem că atomul nostru are un moment magnetic aliniat cu direcţia spinului. Atunci când atomul este plasat într-un câmp magnetic, direcţia spinului se va roti într-un mod bine-definit, ce poate fi calculat cu exactitate din acţiunea lui **U**, pentru a da o nouă stare, să spunem $|\psi\rangle = |\rightarrow\rangle$, după un timp t. Trebuie să considerăm această stare calculată ca pe o parte a realităţii fizice? Ar fi greu de văzut cum anume am putea-o nega. Pentru că $|\psi\rangle$ a fost de la început

preparată pentru posibilitatea că am fi putut s-o măsurăm cu măsurătoarea primitivă la care ne-am referit şi anume cea al cărui spaţiu DA consistă din multipli lui $|\psi>$. În acest caz, este vorba de măsurătoarea de spin în direcţia $\rightarrow$ Sistemul trebuie să ştie să dea răspunsul DA, cu *certitudine* pentru această măsurătoare, pe care nici o altă stare de spin *nu* a atomului, alta decât $|\psi> = |\rightarrow>$, nu o poate garanta.

În practică, există multe feluri de situaţii fizice, diferite de exemplele anterioare, în care o astfel de măsurătoare primitivă ar fi total nepractică. Şi totuşi, regulile standard ale teoriei cuantice permit acestor măsurători să fie, în principiu, efectuate. A nega posibilitatea măsurătorilor de acest tip pentru anumite tipuri prea complicate de vectori de stare ar însemna să schimbăm cadrul teoriei cuantice. Poate că el *trebuie* schimbat (iar în paragraful 6.12 voi avansa sugestii specifice privind această problemă). Sau, cel puţin, este necesar un anumit tip de modificare dacă distincţiile obiective dintre diferite stări cuantice trebuie negate, adică, dacă $|\psi>$ nu este luat, într-un anumit sens fizic clar, ca real în mod *obiectiv*.

Schimbarea 'minimală' sugerată adesea în legătură cu teoria măsurătorii este introducerea a ceea ce se numeşte *reguli de superselecţie* [6], care reduc numărul de tipuri de măsurători efective care pot fi efectuate asupra unui sistem. Nu vreau să intru în detalii, deoarece, după părerea mea, toate aceste sugestii nu au fost dezvoltate până la punctul în care să fie elaborată o imagine coerentă privind problema măsurătorii. Singurul punct pe care vreau să-l subliniez este că o schimbare de această natură, chiar şi minimală, este totuşi o schimbare şi asta ne readuce la punctul de plecare şi anume la observaţia că teoria cuantică are nevoie de o modificare.

În sfârşit, ar trebui să menţionez, poate, că există numeroase alte abordări ale teoriei cuantice care, fără să contrazică predicţiile teoriei convenţionale furnizează 'imagini ale realităţii' care diferă cu mult de cea în care vectorul de stare $|\psi>$ este 'luat în serios' ca reprezentând chiar realitatea. Printre acestea este teoria *ghidului de undă* a prinţului Louis de Broglie (1956) şi al lui David Bohm (1952) o teorie non-locală pentru care există nu numai o funcţie de undă $|\psi>$ ci *şi* un sistem de particule de tip clasic, *ambele* fiind considerate 'reale'. (Vezi, de asemenea, Bohm şi Hilley, 1994.) Alte puncte de vedere implică întregi 'istorii' ale comportamentului posibil, (stimulate de abordarea lui Richard Feynmann din 1948), pentru care punctul de vedere al 'realităţii fizice' diferă întrucâtva de cel furnizat de un vector de stare obişnuit $|\psi>$. Susţinători recenţi ai unei scheme de

acest tip, dar care iau de asemenea în considerare posibilităţile unor măsurători parţiale (Aharonov et.al., 1964), sunt Griffiths, (1984), Omnes (1992), Gell-Mann şi Hartle (1993). Ar fi nepotrivit să mă întind la o discuţie a tuturor acestor alternative aici. Voi spune doar că, deşi aceste proceduri conţin multe puncte de interes considerabil şi o mare cantitate de originalitate stimulatoare, sunt destul de neconvins că problema măsurătorii ar putea fi într-adevăr rezolvată cu ajutorul unor descripţii de aceste tipuri. Desigur, există posibilitatea ca timpul să dovedească că mă înşel.

6.4. Matricea densitate

Mulţi fizicieni susţin că sunt oameni pragmatici şi că, deci, nu sunt interesaţi în problemele privind 'realitatea' lui $|\psi>$. Ei ar spune că tot ce avem nevoie în legătură cu $|\psi>$ este să ne permită să calculăm probabilităţile corespunzătoare comportării fizice viitoare. Deseori, o stare considerată iniţial ca reprezentând o situaţie fizică evoluează în ceva extrem de complicat, unde non-localizarea (datorată mediului înconjurător) devine atât de mare încât nu există nici o posibilitate practică de a vedea efectele de interferenţă cuantică care diferenţiază acea stare de altele similare. Astfel de fizicieni 'pragmatici' vor pretinde, fără îndoială, că nu are sens să susţinem că vectorul de stare particular care a rezultat din această evoluţie are mai multă 'realitate' decât alţii de care este, în practică, imposibil de deosebit. Într-adevăr, vor spune ei, putem foarte bine folosi un *amestec de probabilitate* de stări vectoare pentru a descrie 'realitatea' tot atât de bine ca în cazul folosirii unui vector particular. Ideea este că, dacă aplicarea lui **U** la o anumită stare vectoare care reprezintă starea iniţială a unui sistem ne dă ceva ce, pentru toate scopurile practice (FAPP în formularea lui Bell), este imposibil de distins de un astfel de amestec de probabilitate, atunci amestecul de probabilitate, mai degrabă decât starea vectoare care evoluează cu **U**, este bun pentru descrierea lumii.

Deseori se argumentează că – cel puţin FAPP – procedura **R** poate fi înţeleasă în aceşti termeni. Voi încerca deci să văd dacă este adevărat că (aparentul) paradox **U/R** poate fi rezolvat prin astfel de metode. Să încercăm, însă, mai întâi, să fim puţin mai expliciţi privind procedurile care sunt adoptate în tipul de

abordare FAPP standard pentru explicarea procesului **R** (aparent?).

Cheia pentru aceste proceduri este un obiect matematic numit *matricea densitate*. Matricea densitate este un concept important în teoria cuantică, aflat la baza celor mai multe descrieri matematice standard ale procesului de măsurare. Şi în abordarea mea, mai puţin convenţională, locul matricei densitate este important, în special în relaţie cu procedurile FAPP standard. Din aceste motive voi fi nevoit, din păcate, să intru ceva mai mult în formalismul matematic al teoriei cuantice decât până acum. Sper că cititorul neiniţiat nu se va lăsa speriat. Chiar dacă nu înţelegeţi cu totul argumentele care urmează, cred că va fi de ajutor să frunzăriţi argumentele matematice pe măsură ce ele intervin şi să le surprindeţi o parte din savoare. Acest lucru va fi de o valoare considerabilă pentru înţelegerea argumentaţiilor ulterioare şi a subtilităţilor implicate în discuţia privind motivul pentru care *avem nevoie* de o teorie cuantică modificată!

Putem privi matricea densitate ca reprezentând amestecul de probabilitate al unui număr de posibili vectori de stare *alternativi*, mai degrabă decât un singur vector de stare. Printr-un amestec probabilist înţelegem doar că există o anumită nedeterminare privind starea actuală a sistemului, în care fiecărei stări vectoare alternative îi este asociată o probabilitate. Acestea sunt probabilităţi în sensul clasic, descrise prin numere reale. În cazul unei matrice densitate există însă o confuzie (deliberată) în această descriere, între probabilităţile *clasice*, care apar în acest amestec de probabilitate şi *probabilităţile cuantice* care ar rezulta din procedura **R**. Ideea este că nu putem distinge operaţional între cele două, astfel încât o descriere matematică matricea densitate — care *nu* distinge între ele este operaţional corespunzătoare.

Dar care este această descriere matematică? Nu vreau să intru în prea multe detalii, însă o trecere în revistă a conceptelor de bază ar fi utilă. Ideea matricei-densitate este, de fapt, foarte elegantă[1] Mai întâi, în locul fiecărei stări individuale $|\psi\rangle$, folosim un obiect scris $|\psi\rangle\langle\psi|$

Ce înseamnă asta? Definiţia matematică precisă nu este importantă în

[1] Ea a fost introdusă în 1932 de faimosul matematician maghiaro–american John von Neumann cel care, în plus, a fost singura persoană responsabilă pentru teoria care a stat la baza construirii calculatoarelor electronice (pornind de la ideile lui Alan Turing). Teoria jocurilor, la care m–am referit în nota 9 de la capitolul 3 a fost, de asemenea, creaţia lui von Neumann. În plus, el a fost prima persoană care a distins procedurile pe care le–am numit **U** şi **R**.

momentul acesta. Să spunem că expresia reprezintă un fel de 'produs' (o formă a produsului tensorial la care ne-am referit în paragraful 5.15) între vectorul de stare $|\psi\rangle$ şi conjugatul complex, scris $\langle\psi|$. Considerăm că $|\psi\rangle$ este un vector de stare *normalizat* ($\langle\psi|\psi\rangle = 1$) şi deci că expresia $|\psi\rangle\langle\psi|$ este unic determinată de starea fizică pe care vectorul $|\psi\rangle$ o reprezintă (independent de factorul de fază $|\psi\rangle \mapsto e^{i\theta}|\psi\rangle$, discutat în paragraful 5.10). În terminologia Dirac, vectorul original $|\psi\rangle$ este numit vector 'ket', iar $\langle\psi|$, vectorul 'bra' corespunzător. Împreună un bra-vector $\langle\psi|$ şi un ket-vector $|\phi\rangle$ se pot combina pentru a forma *produsul scalar* ('bracket'): $\langle\psi|\phi\rangle$ o notaţie pe care cititorul o va recunoaşte din paragraful 5.12. Acest produs scalar este doar un număr complex obişnuit, pe când produsul tensorial $|\psi\rangle\langle\phi|$ care apare cu o matrice densitate este un 'obiect' matematic mult mai complex un element al unui anumit spaţiu vectorial.

Există o operaţie matematică specifică, numită 'a lua urma'; ea ne permite trecerea de la acest 'obiect' matematic, la un număr complex obişnuit. Pentru o expresie $|\psi\rangle\langle\phi|$ această operaţie revine la a inversa ordinea termenilor în produsul scalar:

$$trace\ (|\psi\rangle\langle\phi|) = \langle\phi|\psi\rangle$$

Pe când, pentru o sumă de termeni, 'urma' acţionează liniar; de exemplu:

$$trace\ (z|\psi\rangle\langle\phi| + w|\alpha\rangle\langle\beta|) = z\langle\phi|\psi\rangle + w\langle\beta|\alpha\rangle$$

Nu voi intra în detalii privind proprietăţile matematice ale obiectelor de tip $\langle\psi|$ şi $|\psi\rangle\langle\phi|$, însă unele dintre ele sunt importante. În primul rând, produsul $|\psi\rangle\langle\phi|$ satisface exact aceleaşi legi algebrice pe care le-am văzut în paragraful 5.15 $|\psi\rangle|\phi\rangle$ (exceptând-o pe ultima, care, de altfel, nu este relevantă aici):

$$(z|\psi\rangle)\langle\phi| = z(|\psi\rangle\langle\phi|) = |\psi\rangle(z\langle\phi|)$$
$$(|\psi\rangle+|\chi\rangle)\langle\phi| = |\psi\rangle\langle\phi| + |\chi\rangle\langle\phi|$$
$$|\psi\rangle(\langle\phi|+\langle\chi|) = |\psi\rangle\langle\phi| + |\psi\rangle\langle\chi|$$

Trebuie să observăm că vectorul bra $z\langle\psi|$ este conjugatul complex al lui ket $z|\psi\rangle$, z fiind conjugatul complex al numărului complex z, iar $\langle\psi|\langle\chi|$ este conjugatul complex lui $|\psi\rangle+|\chi\rangle$.

Să presupunem că vrem să descriem matricea densitate care reprezintă un anumit amestec de probabilitate al stărilor normalizate, să spunem $|\alpha\rangle$ şi $|\beta\rangle$,

cu probabilităţile respective a şi b. Matricea densitate va fi

$$\boldsymbol{D} = a|\alpha\rangle\langle\alpha| + b|\beta\rangle\langle\beta|$$

Pentru trei stări normalizate, $|\alpha\rangle$, $|\beta\rangle$, $|\gamma\rangle$ cu probabilităţile respective a, b, c, avem $\boldsymbol{D} = a|\alpha\rangle\langle\alpha| + b|\beta\rangle\langle\beta| + c|\gamma\rangle\langle\gamma|$ şi aşa mai departe. Din faptul că probabilităţile pentru toate alternativele trebuie să se adune pentru a da unu, se poate deduce următoarea proprietate importantă, valabilă pentru orice matrice densitate:

$$trace\ (\boldsymbol{D}) = 1$$

Cum putem folosi o matrice densitate pentru a calcula probabilităţile care apar dintr-o anumită măsurătoare? Să considerăm mai întâi cazul unei măsurători primitive. Întrebăm sistemul dacă este într-o stare fizică $|\psi\rangle$ (DA) sau într-una ortogonală pe $|\psi\rangle$ (NU). Măsurătoarea însăşi este reprezentată printr-un obiect matematic (numit *proiector*) similar cu matricea densitate:

$$\boldsymbol{E} = |\psi\rangle\langle\psi|$$

Probabilitatea p de a obţine DA este $p = trace\ (\boldsymbol{DE})$, unde produsul $\boldsymbol{DE}$ este el însuşi un 'obiect' de tipul matricei densitate. De exemplu, pentru suma de doi termeni de mai sus, $\boldsymbol{D} = a|\alpha\rangle\langle\alpha| + b|\beta\rangle\langle\beta|$, avem:

$$\begin{aligned}\boldsymbol{DE} &= (a|\alpha\rangle\langle\alpha| + b|\beta\rangle\langle\beta|)|\psi\rangle\langle\psi| = \\ &= a|\alpha\rangle\langle\alpha|\psi\rangle\langle\psi| + b|\beta\rangle\langle\beta|\psi\rangle\langle\psi| = \\ &= (a\langle\alpha|\psi\rangle)|\alpha\rangle\langle\psi| + (b\langle\beta|\psi\rangle)|\beta\rangle\langle\psi|\end{aligned}$$

Termenii $\langle\alpha|\psi\rangle$ şi $\langle\beta|\psi\rangle$ pot 'comuta' cu alte expresii, deoarece sunt doar numere, însă trebuie să fim atenţi cu ordinea 'obiectelor' de tip $|\alpha\rangle$ sau $\langle\psi|$. Amintiţi-vă că $z\bar{z}=|z|^2$ aşa ca în paragraful 5.9, deducem:

$$\begin{aligned}trace\ (\boldsymbol{DE}) &= (a\langle\alpha|\psi\rangle)\langle\psi|\alpha\rangle + (b\langle\beta|\psi\rangle)\langle\psi|\beta\rangle = \\ &= a|\langle\alpha|\psi\rangle|^2 + b|\langle\beta|\psi\rangle|^2\end{aligned}$$

Amintiţi-vă că $|\langle\alpha|\psi\rangle|^2$ şi $|\langle\beta|\psi\rangle|^2$ sunt probabilităţi *cuantice* (vezi paragraful 5.13) pentru rezultatele respective $|\alpha\rangle$ şi $|\beta\rangle$ pe când a şi b furnizează contribuţiile *clasice* la probabilitatea totală. Astfel, probabilităţile cuantice şi probabilităţile clasice sunt amestecate în expresia finală.

Pentru măsurătoarea mai generală de tip da/nu, discuţia este în principiu aceeaşi, cu excepţia faptului că în locul lui $\boldsymbol{E}$, definit mai sus, vom folosi un proiector mai general, ca de pildă:

$$\boldsymbol{E} = |\psi\rangle\langle\psi| + |\phi\rangle\langle\phi| + \quad + |\chi\rangle\langle\chi|$$

unde $|\psi\rangle$, $|\phi\rangle$, $|\chi\rangle$ sunt stări mutuale ortogonale normalizate în spaţiul

stărilor DA din spaţiul Hilbert. Avem proprietatea generală $\boldsymbol{E}^2 = \boldsymbol{E}$ ce caracterizează un proiector. Probabilitatea de DA pentru măsurătorile definite de proiectorul $\boldsymbol{E}$ asupra sistemului cu matricea de densitate $\boldsymbol{D}$ este *trace(**DE**)*, exact ca înainte.

Să observăm faptul (important) că probabilitatea cerută poate fi calculată dacă ştim matricea densitate şi proiectorul care descrie măsurătoarea. Nu avem nevoie să ştim felul particular în care matricea densitate a fost alcătuită din stările particulare. Probabilitatea totală rezultă automat ca o combinaţie potrivită de probabilităţi clasice şi cuantice, fără să trebuiască să ne îngrijorăm prea mult cât din probabilitatea rezultantă vine din fiecare parte.

Să examinăm mai îndeaproape acest mod curios în care probabilităţile cuantice şi clasice sunt întreţesute în matricea densitate. Să presupunem de exemplu, că avem o particulă de spin ½ şi că nu ştim dacă starea (normalizată) de spin este $|\uparrow>$ sau $|\downarrow>$. Astfel, considerând probabilităţile respective ½ şi ½, matricea densitate devine:

$$\boldsymbol{D} = ½|\uparrow><\uparrow| + ½|\downarrow><\downarrow|$$

Rezultă, dintr-un calcul simplu, că exact aceeaşi matrice densitate D va apărea în cazul unui amestec de probabilitate ½, ½ , a oricăror alte posibilităţi ortogonale, de exemplu stările normalizate $|\rightarrow>$ şi $|\leftarrow>$ (unde $|\rightarrow> = (|\uparrow> + |\downarrow>)/\sqrt{2}$ şi $|\leftarrow> = (|\uparrow> - |\downarrow>)/\sqrt{2}$):

$$\boldsymbol{D} = ½|\rightarrow><\rightarrow| + ½|\leftarrow><\leftarrow| \,.$$

Să presupunem că alegem să măsurăm spinul particulei în direcţia 'sus', astfel încât proiectorul relevant este $\boldsymbol{E} = |\uparrow><\uparrow|$ Atunci găsim, pentru probabilitatea lui DA, conform primei descrieri,

$$\textit{trace}(\boldsymbol{DE}) = ½|<\uparrow|\uparrow>|^2 + ½|<\downarrow|\uparrow>|^2 = ½\times1^2 + ½\times0^2 = ½$$

Unde folosim $<\uparrow|\uparrow> = 1$ şi $<\downarrow|\uparrow> = 0$ (stările fiind normalizate şi ortogonale); iar, conform celei de-a doua

$$\textit{trace}(\boldsymbol{DE}) = ½|<\rightarrow|\uparrow>|^2 + ½|<\leftarrow|\uparrow>|^2 =$$
$$= ½\times(1/\sqrt{2})^2 + ½\times(1/\sqrt{2})^2 = ¼ + ¼ = ½ \,,$$

unde acum stările dreapta/stânga $|\rightarrow>$, $|\leftarrow>$ nu sunt nici ortogonale nici paralele cu starea măsurată $|\uparrow>$ şi, de fapt, $|<\rightarrow|\uparrow>| = |<\leftarrow|\uparrow>| = 1/\sqrt{2}$

Cu toate că probabilităţile rezultante sunt egale, (aşa cum trebuie să fie, deoarece matricea densitate este aceeaşi) interpretarea fizică a acestor două descrieri este destul de diferită. Acceptând că 'realitatea' fizică a unei situaţii

trebuie descrisă printr-un vector de stare diferit; există însă o anumită nedeterminare clasică privind care anume poate fi vectorul de stare actual. În prima dintre cele două descrieri de mai înainte, starea este fie $|\uparrow>$ fie $|\downarrow>$, dar nu ştim care anume. În a doua, este fie $|\rightarrow>$ sau $|\leftarrow>$ dar nu ştim care anume. În prima descriere, dacă efectuăm o măsurătoare întrebând dacă starea este $|\uparrow>$, e o problemă de probabilităţi clasice: există, într-adevăr o probabilitate ½ ca starea să fie $|\uparrow>$ şi asta e tot. În a doua descriere, unde punem aceeaşi întrebare, măsurătoarea întâlneşte amestecul de probabilitate $|\rightarrow>$ şi $|\leftarrow>$, iar fiecare dintre stări are o contribuţie clasică de ½ ori o contribuţie cuantică de ½ cu totalul ¼ + ¼ = ½ Vedem astfel că matricea densitate contribuie pentru a ne da probabilitatea corectă, indiferent cum este considerată această probabilitate în raport cu părţile ei clasică şi cuantică.

Exemplul de mai sus este special pentru că matricea densitate posedă 'valori proprii degenerate' (faptul că, aici, valorile celor două probabilităţi clasice sunt egale) ceea ce ne permite să avem mai mult decât o descriere în termenii amestecurilor de probabilitate al alternativelor ortogonale. Totuşi, acesta nu este un punct esenţial pentru discuţia noastră (l-am menţionat aici doar pentru a-i linişti pe experţi). Putem permite stărilor alternative dintr-o mixtură de probabilitate să conţină mai multe stări decât o singură mulţime de alternative mutual ortogonale. De exemplu, în situaţia anterioară, putem avea mixturi de probabilitate complicate pentru multe direcţii diferite posibile de spin. Rezultă că pentru *orice* matrice densitate — nu doar pentru cele cu valori proprii degenerate — există multe feluri complet diferite de a reprezenta aceeaşi matrice densitate ca pe un amestec de probabilitate al stărilor alternative.

6.5. Matrice densitate pentru perechi EPR

Să examinăm acum un tip de situaţie pentru care descrierea cu ajutorul unei matrice densitate este în mod particular potrivită — şi care, totuşi, ne indică un aspect aproape paradoxal al acestei interpretări, în legătură cu efectele EPR şi non-localizarea cuantică. Să considerăm situaţia fizică discutată în paragraful 5.17, în care o particulă de spin 0 (în starea $|\Omega>$) se scindează în două particule de spin ½ care se deplasează spre dreapta şi spre stânga, la mare distanţă una de alta.

Starea lor combinată (non-separabilă) de spin este:

$$|\Omega> = |L\uparrow>|R\downarrow> - |L\downarrow>|R\uparrow>$$

Să presupunem că spinul particulei din partea dreaptă este examinat de un aparat de măsură, în timp ce particula din stânga a ajuns atât de departe încât observatorul nu mai are acces la ea. Cum va descrie observatorul starea de spin a particulei din partea dreaptă?

Matricea densitate în acest caz este

$$\boldsymbol{D} = \tfrac{1}{2}|R\uparrow><R\uparrow| + \tfrac{1}{2}|R\downarrow><R\downarrow|$$

Ne putem imagina că un alt observator — un coleg aflat la mare distanţă a ales să măsoare spinul particulei din partea stângă într-o direcţie verticală. El nu are nici un mod de a spune ce rezultat a obţinut colegul lui pentru această măsurătoare de spin. Ceea ce ştie, însă, este că dacă colegul lui al obţinut rezultatul $|L\uparrow>$, atunci particula sa trebuie să fie în starea $|R\downarrow>$. Mai ştie de asemenea (din legile standard ale teoriei cuantice) că probabilitatea ca imaginarul său coleg să fi obţinut ca rezultat $|L\uparrow>$ este egală cu probabilitatea de a obţine $|L\downarrow>$ Astfel, concluzia lui va fi că starea propriei sale particule este un amestec de probabilitate similar (adică, cu probabilităţile respective ½, ½) pentru cele două alternative $|R\uparrow>$ şi $|R\downarrow>$, astfel că matricea sa densitate trebuie într-adevăr să fie $\boldsymbol{D}$, în forma precedentă.

El ar putea, totuşi, să-şi imagineze că măsurătoarea colegului său a fost efectuată în direcţia dreapta-stânga şi nu sus-jos. Exact acelaşi tip de raţionament (folosind acum descrierea alternativă $\boldsymbol{D} = |L\leftarrow>|R\rightarrow> - |L\rightarrow>|R\leftarrow>$) îl va duce la concluzia că starea de spin a particulei sale este un amestec de probabilitate egală pentru cele două direcţii, dreapta şi stânga, ceea ce îi furnizează matricea densitate

$$\boldsymbol{D} = \tfrac{1}{2}|R\rightarrow><R\rightarrow| + \tfrac{1}{2}|R\leftarrow><R\leftarrow|$$

Aşa cum am spus mai înainte, aceasta este exact aceeaşi matrice densitate pe care o aveam înainte, însă *interpretarea* ei ca un amestec de probabilitate al unor stări alternative este diferit! Nu contează ce interpretare adoptă observatorul. Matricea densitate îi furnizează toate informaţiile existente pentru calcularea probabilităţilor rezultatului măsurătorilor de spin asupra particulei din dreapta. Mai mult, deoarece colegul său este un coleg imaginar, observatorul nostru nu are nevoie să considere că asupra particulei din stânga s-a efectuat vreo măsurătoare. Într-adevăr, am putea presupune că 'starea actuală' a particulei din dreapta este

determinată mai corect cu ajutorul matricei densitate decât cu ajutorul oricărui vector de stare specific.

Consideraţii generale de acest tip conduc uneori la ideea că matricele densitate furnizează o descriere mai potrivită a 'realităţii' cuantice în anumite circumstanţe decât o fac vectorii de stare. Totuşi, acest lucru nu ne va furniza un punct de vedere comprehensiv în situaţii ca cea menţionată. Pentru că nu există nimic care să împiedice posibilitatea ca acel coleg imaginar să devină unul real, iar cei doi obsesvatori de să-şi comunice unul altuia rezultatele. Corelaţiile dintre măsurătorile unuia dintre observatori şi ale celuilalt nu pot fi explicate în funcţie de matricele densitate separate pentru cele două particule din dreapta şi din stânga separat. Pentru aceasta am avea nevoie de întreaga stare non-localizată furnizată de expresia pentru vectorul de stare actual, $|\Omega>$, ca mai înainte.

De exemplu, dacă ambii observatori aleg să măsoare spinul particulelor în direcţia sus/jos, atunci ei trebuie în mod necesar să obţină răspunsuri opuse pentru rezulatatele măsurătorilor lor. Matricele densitate individuale pentru cele două particule nu vor furniza această informaţie. Chiar mai serios, teorema lui Bell (paragraful 5.4) arată că nu există un model clasic de tip local (de tipul 'şosetelor lui Bertlmann') pentru starea non-separabilă a perchii de paticule înainte de măsurătoare. Vezi **MN**, Capitolul 6, nota 14, pentru o demonstraţie simplă a acestui fapt — datorată în mod esenţial lui Stapp, (1979, 1993) — în cazul în care unul dintre observatori alege să măsoare spinul particule sale fie sus/jos fie dreapta/stânga, pe când celălalt alege în una dintre cele două direcţii la 45° faţă de cele două. Dacă înlocuim cele două particule de spin ½ de două particule de spin 3/2, atunci dodecaedrele magice din paragraful 5.3 arată acest gen de lucruri chiar mai convingător, deoarece în cazul lor nu este nevoie de probabilităţi.

Acest lucru arată că descrierea cu ajutorul matricei densitate poate fi adecvată pentru descrierea 'realităţii' situaţiei doar dacă există vreun motiv *în principiu* pentru care măsurătorile asupra celor două părţi ale sistemului nu pot fi efectuate şi comparate — ceea ce nu pare să fie cazul, în situaţiile normale. În cele anormale — ca cea imaginată de Stephen Hawking (1982) în care una dintre particulele unei perechi EPR este capturată în interiorul unei găuri negre — poate exista un motiv foarte serios pentru o descriere cu ajutorul matricei densitate la nivel fundamental (aşa cum argumentează Hawking). Dar acest lucru va constitui, el însuşi, o schimbare în însuşi cadrul teoriei cuantice. Iar fără o astfel de

schimbare, rolul esenţial al matricei densitate este mai degrabă FAPP decât fundamental — deşi rolul său nu este deloc neglijabil.

6.6. O explicaţie FAPP a lui R?

Să vedem acum cum îşi joacă matricele densitate rolul în abordarea standard — FAPP — a modului în care 'pare' să aibă loc procesul **R**. Ideea este că un sistem cuantic şi un aparat de măsură, împreună cu mediul în care se află — toate fiind presupuse a evolua conform lui **U** — se vor comporta *ca şi când* **R** a avut loc ori de câte ori efectele măsurătorii devin inextricabil neseparabile de mediu.

Sistemul cuantic este iniţial considerat izolat de mediul său, însă în cursul 'măsurării' produce efecte considerabile în aparatul de măsură care vor implica în curând non-separabilitatea cu părţi considerabile şi din ce în ce mai mari ale mediului. În acest stadiu, imaginea devine similară cu cea a situaţiei EPR discutată în secţiunea precedentă. Sistemul cuantic, împreună cu aparatul de măsură joacă rolul particulei din partea dreaptă, pe când mediul perturbat joacă rolul particulei din partea stângă. Un fizician care-şi propune să examineze aparatul de măsură va juca un rol similar cu cel al observatorului din discuţia precedentă, care examinează particula din partea dreaptă. Observatorul nu are nici un fel de acces la vreo măsurătoare care ar putea fi efectuată asupra particulei din partea stângă; în mod asemănător, fizicianul nu are acces la modul detaliat în care mediul poate fi perturbat de către aparatul de măsură. Mediul constă dintr-un număr enorm de particule cu mişcare aleatoare şi putem considera că informaţia detaliată conţinută în modul precis în care particulele mediului au fost perturbate va fi, în practică, definitiv pierdută pentru fizicianul nostru. Acest lucru este similar faptului că orice informaţie privind spinul particulei din partea stângă, din exemplul de mai sus, este inaccesibilă observatorului din partea dreaptă. Ca şi în cazul particulei din partea dreaptă, starea aparatului de măsură este descrisă corespunzător de o matrice densitate şi nu de o stare cuantică pură. Acest amestec de probabilitate furnizează alternativele ponderate de probabilităţi pe care procedura **R** ni le-ar fi dat — cel puţin FAPP. Sau, cel puţin, aşa se desfăşoară argumentaţia standard.

Să considerăm un exemplu. Să presupunem că un foton este emis de o

anumită sursă, în direcţia unui detector. Între sursă şi detector avem o oglindă parţial argintată. După ce întâlneşte oglinda, starea fotonului se află în superpoziţia

$$w|\alpha\rangle + z|\beta\rangle ,$$

unde starea transmisă $|\alpha\rangle$ va activa detectorul (DA) dar starea reflectată $|\beta\rangle$ îl va lăsa neperturbat (NU). Presupun aici că toate stările sunt normalizate, astfel încât, conform procedurii **R**, vom obţine

probabilitatea lui DA = $|w|^2$; probabilitatea lui NU = $|z|^2$

Pentru o oglindă semiargintată (ca în exemplul iniţial considerat în paragraful 5.7, unde $|\alpha\rangle$ şi $|\beta\rangle$ vor fi stările $|B\rangle$ şi respectiv $i|C\rangle$) aceste două posibilităţi sunt, fiecare ½ şi avem $|w| = |z| = 1/\sqrt{2}$

Detectorul este iniţial în starea $|\Psi\rangle$, care evoluează în $|\Psi_D\rangle$ (DA) dacă absoarbe fotonul (în starea $|\alpha\rangle$) şi care evoluează în $|\Psi_N\rangle$ (NU) dacă nu absoarbe fotonul (în starea $|\beta\rangle$). Dacă mediul ar putea fi ignorat, atunci starea totală în acest stadiu va fi de forma $w|\Psi_D\rangle + z|\Psi_N\rangle|\beta\rangle$ (toate stările fiind presupuse normalizate); să presupunem însă că detectorul, fiind un obiect macroscopic, ajunge foarte repede să fie implicat în interacţiuni cu mediul înconjurător — şi putem presupune că fotonul (în mod iniţial în starea $|\beta\rangle$) este absorbit de către peretele laboratorului pentru a deveni, de asemenea, o parte a mediului. Ca şi înainte, în funcţie de detectarea sau non-detectarea fotonului, aparatul de măsură se va afla în starea $|\Psi_D\rangle$ sau $|\Psi_N\rangle$, respectiv, dar stabilirea într-una din aceste stări provoacă perturbarea stării mediului care devine $|\Phi_D\rangle$ pentru $|\Psi_D\rangle$, sau $|\Phi_N\rangle$ pentru $|\Psi_N\rangle$ (din nou presupunem stările normalizate, dar nu în mod necesar ortogonale). Starea totală va fi $w|\Phi_D\rangle|\Psi_D\rangle + z|\Phi_N\rangle|\Psi_N\rangle$

Până aici, fizicianul nu este implicat, dar iată-l acum pe punctul de a examina detectorul pentru a vedea dacă a înregistrat DA sau NU. Cum trebuie să vedem starea cuantică a detectoului chiar înainte de a fi examinat? Ca şi în cazul observatorului care măsura spinul particulei din partea dreaptă, din discuţia precedentă, este adecvată folosirea matricei densitate. Putem presupune că nici o măsurătoare nu este efectuată în mod actual asupra mediului pentru a stabili dacă starea lui este $|\Phi_D\rangle$ sau $|\Phi_N\rangle$, asemănător cu cazul particulei din partea stângă a perechii EPR descrisă anterior. Corespunzător, matricea densitate furnizează într-adevăr o descriere cuantică adecvată a detectorului.

Ce este această matrice densitate? Tipul standard de argument [7] (bazat pe un mod perticular de a modela mediul şi, de asemenea, pe anumite presupoziţii

incomplet justificate, ca de pildă iportanţa tipului de corelaţii EPR) duce la concluzia că această matrice densitate va trebui să se apropie rapid şi foarte mult de forma:

$$\boldsymbol{D} = a|\Psi_D\rangle\langle\Psi_D| + b|\Psi_N\rangle\langle\Psi_N|, \text{ unde } a = |w|^2 \text{ şi } b = |z|^2$$

Această matrice densitate poate fi interpretată ca reprezentând un amestec de probabilitate a detectorului înregistrând DA cu probabilitatea $|w|^2$ şi înregistrând NU cu probabilitatea $|z|^2$. Acest lucru este exact ceea ce ne-ar fi spus procesul **R** că vom găsi ca rezultat al experimentului — sau nu este aşa?

Trebuie să avem grijă cu această concluzie. Matricea densitate $\boldsymbol{D}$ va permite într-adevăr fizicianului să calculeze probabilităţile de care are nevoie, dacă-i este permis să *presupună* că alternativa pe care o are la dispoziţie este doar cea conform căreia starea detectorului este *sau* $|\Psi_D\rangle$ *sau* $|\Psi_N\rangle$. Însă această presupunere nu este, în nici un fel, o consecinţă a discuţiei noastre. Amintiţi-vă din discuţia precedentă că matricele densitate pot avea mai multe interpretări *alternative* ca amestecuri de probabilitate. În cazul specific al unei oglinzi semiargintate, obţinem o matrice densitate de exact aceeaşi formă ca în cazul particulei de spin ½

$$\boldsymbol{D} = ½|\Psi_D\rangle\langle\Psi_D| + ½|\Psi_N\rangle\langle\Psi_N|$$

Ceea ce poate fi exprimat ca

$$\boldsymbol{D} = ½|\Psi_P\rangle\langle\Psi_P| + ½|\Psi_Q\rangle\langle\Psi_Q|$$

unde $|\Psi_P\rangle$ şi $|\Psi_Q\rangle$ sunt două stări posibile diferite ortogonale pentru detector — stări care vor fi destul de absurde din punctul de vedere al fizicii clasice, ca:

$$|\Psi_P\rangle = (|\Psi_D\rangle + |\Psi_N\rangle)/\sqrt{2} \text{ şi } |\Psi_Q\rangle = (|\Psi_D\rangle - |\Psi_N\rangle)/\sqrt{2}$$

Faptul că fizicianul consideră că starea detectorului său este descrisă de matricea densitate $\boldsymbol{D}$ nu explică în nici un fel de ce găseşte el că detectorul este întotdeauna fie într-o stare DA (dată de $|\Psi_D\rangle$) fie într-o stare NU (dată de $|\Psi_N\rangle$). Pentru că exact aceeaşi matrice densitate ar fi dată dacă starea ar fi o combinaţie de pondere egală de absurdităţi clasice $|\Psi_P\rangle$ şi $|\Psi_Q\rangle$ (care descriu, respectiv, superpoziţiile liniare "DA plus NU" şi "DA minus NU")!

Pentru a sublinia absurditatea fizică a stărilor precum $|\Psi_P\rangle$ şi $|\Psi_Q\rangle$ pentru un detector macroscopic, să considerăm cazul unui 'aparat de măsură' alcătuit dintr-o cutie cu o pisică înăuntru, unde există, de asemenea, un dispozitiv care omoară pisica dacă detectorul primeşte un foton (în starea $|\alpha\rangle$) dar nu şi în caz contrar (fotonul în starea $|\beta\rangle$). Adică povestea *pisicii lui Schrödinger* (vezi paragraful 5.1, figura 6.3). Răspunsul DA va fi prezentat în forma 'pisicii moarte'

iar răspunsul NU reprezintă 'pisica vie'. Totuşi, simplul fapt de a şti că matricea densitate are forma unui amestec egal al acestor două stări, *nu* ne spune dacă pisica este moartă sau vie (cu probabilităţi egale), deoarece poate fi le fel de bine "moartă plus vie" sau "moartă minus vie", tot cu probabilităţi egale! Matricea densitate singură *nu* ne spune că aceste două posibilităţi absurde nu vor apărea niciodată în lumea pe care o cunoaştem. Ca şi în cazul tipului de interpretare *many worlds*, părem obligaţi, din nou, să luăm în considerare ce tip de stări este în stare observatorul conştient (aici, fizicianul nostru) să perceapă. De ce ar fi, până la urmă, starea "mort plus viu" ceva imposibil de perceput pentru un observator conştient extern[2] ?

Unii ar putea răspunde că 'măsurătoarea' pe care fizicianul nostru era pe punctul de a o efectua cu detectorul n-a făcut, până la urmă, decât să determine dacă detectorul înregistrează DA sau NU — adică, să stabilească, în exemplul nostru, dacă pisica este moartă sau vie (este cazul observatorului din secţiunea precedentă care determina dacă spinul particulei din partea dreaptă este 'sus' sau 'jos'). Pentru această măsurătoare, matricea densitate ne dă într-adevăr probabilităţile corecte, indiferent în ce fel alegem să le reprezentăm. Totuşi, întrebarea se impune: de ce faptul că *privim* pisica produce o măsurătoare de acest tip. Nu există nimic în evoluţia **U** a unui sistem cuantic care să ne spună că în actul de a 'privi' şi, în consecinţă, de a *percepe* un sistem cuantic ne este interzis să întâlnim ceva de tipul combinaţiei „pisică vie plus pisică moartă" În acest fel ne-am întors de unde am plecat. Ce este conştienţa? Cum este creierul construit? Motivul pentru care a fost adoptat, de la început, punctul de vedere FAPP, a fost evitarea unor consideraţii de acest tip!

Din nou, unii ar putea argumenta că am ales un caz special nereprezentativ în exemplul nostru, în care cele două probabilităţi ½ şi ½ sunt *egale* (cazul valorilor proprii degenerate). Doar în astfel de situaţii matricea densitate poate fi reprezentată în mai multe feluri ca un amestec ponderat de probabilităţi de alternative mutual ortogonale. Aceasta nu este, totuşi, o restricţie importantă, deoarece ortogonalitatea alternativelor nu este o cerinţă pentru interpretarea unei

---

[2] Ar trebui considerată, desigur, problema conştiinţei proprii a pisicii! Acest aspect a fost subliniat de o versiune a paradoxului pisicii lui Schrödinger datorată lui Eugene P. Wigner (1961). „Prietenul lui Wigner" suferă soarta umilitoare a pisicii lui Schrödinger, dar este în întregime conştient în fiecare dintre stările lui suprapuse!

matrice densitate ca un amestec ponderat de probabilităţi. De fapt, într-o lucrare recentă, Hughston et al. (1993) au arătat că în situaţii ca cea considerată, în care matricea densitate apare din cauză că sistemul considerat este non-separabil de un alt sistem separat, atunci în orice fel am alege să reprezentăm matricea densitate ca un amestec de proababilităţi de stări alternative, există întotdeauna o măsurătoare care poate fi efectuată asupra acelui sistem separat care dă un mod particular de a reprezenta matricea densitate. În orice caz, deoarece ambiguitatea este prezentă în cazul în care probabilităţile sunt egale, atunci acest lucru ne spune că descrierea cu ajutorul matricei densitate nu este suficientă pentru descrierea stărilor actuale alternative ale detectorului.

Concluzia tuturor acestora este că simpla cunoaştere a faptului că matricea densitate este un **D** oarecare *nu* ne spune dacă sistemul este un amestec de probabilităţi a unor seturi de stări particulare care dau naştere acestui **D** particular. Există aproape întotdeauna numeroase moduri diferite de a obţine un acelaşi **D**, cele mai multe dintre ele fiind 'absurde' pentru punctul de vedere al simţului comun. Mai mult, acest tip de ambiguitate se păstrează pentru orice matrice densitate posibilă.

Discuţia standard nu progresează îndeobşte dincolo de încercările de a arăta că **D** este diagonal, ceea ce înseamnă, de fapt, că poate fi exprimat ca un amestec de probabilitate ponderat al unor stări alternative mutual ortogonale — sau, mai degrabă, că poate fi astfel exprimat, atunci când alternativele sunt alternativele clasice care ne interesează (fără această condiţie, toate matricele densitate ar fi diagonale!). Dar am văzut că simplul fapt că matricea densitate poste fi exprimată în acest fel nu ne spune, în sine, că detectorii nu vor fi percepuţi în superpoziţii cuantice absurde de DA şi NU în acelaşi timp.

Astfel, contrar a ceea ce se pretinde frecvent, argumentul standard *nu* explică cum anume are 'iluzia' lui **R** loc ca un tip de descriere aproximativă a evoluţiei **U**, atunci când influenţa mediului devine importantă. Ceea ce *arată* această argumentaţie este că procedura **R** poate coexista paşnic, în anumite circumstanţe, cu evoluţia **U**. Avem încă nevoie de **R** ca parte a teoriei cuantice care e separată de evoluţia **U** (cel puţin în absenţa unei anumite teorii care să ne spună ce fel de stări pot percepe fiinţele conştiente).

Acest lucru este important în sine pentru consistenţa generală a teoriei cuantice. Însă este de asemenea important să realizăm că această consistenţă şi

coexistenţă are un statut FAPP şi nu unul riguros. Amintiţi-vă, din discuţia de la sfârşitul secţiunii precedente, că descrierea cu ajutorul matricei densitate a particulei din partea dreaptă era adecvată doar în absenţa unei comparaţii posibile între măsurătoarele care pot fi efectuate asupra ambelor particule. Pentru asta este necesară întreaga stare, cu superpoziţiile sale *cuantice* şi nu doar cele poderate de probabilităţi. În acelaşi fel, descrierea cu ajutorul matricei densitate a detectorului este adecvată doar dacă detaliile fine ale mediului nu pot fi măsurate şi comparate cu rezultatele observaţiilor efectuate de observator asupra detectorului. **R** poate coexista cu **U** doar dacă detaliile fine ale mediului sunt imune la măsurare; iar efectele de interferenţă cuantică subtilă care (conform teoriei cuantice) se află ascunse în imensa complicaţie a descrierii detaliate a mediului nu pot fi niciodată observate.

Este clar că argumentul standard conţine o oarecare cantitate de adevăr; totuşi, el nu poate nicidecum constitui răspunsul întreg. Cum putem fi noi siguri că efectele unor astfel de fenomene de interferenţă nu vor fi descoperite într-un viitor cu o tehnologie mai avansată? Ar trebui să avem o nouă regulă fizică care să ne spună că anumite experimente care nu pot fi în prezent efectuate *în practică*, nu vor putea fi niciodată efectuate *în principiu*. Conform unei astfel de reguli, ar trebui să existe un anumit nivel al acţiunii fizice la care trebuie să fie, în principiu, imposibil să regăsim aceste efecte de interferenţă. Pare că un *nou* fenomen fizic ar trebui să intervină pentru ca superpoziţiile cu pondere numere complexe de la nivel cuantic să devină alternative fizice la nivel clasic în mod *actual* şi nu doar FAPP Punctul de vedere FAPP, aşa cum apare, nu ne oferă nici un fel de imagine a realităţii fizice. În consecinţă, FAPP trebuie să fie doar un prag de impas al teoriei fizice actuale, un punct de inflexiune valoros a cărui importanţă se va dovedi în cele ce urmează.

6.7. Explică FAPP regula modulului pătrat?

În toate cele trei secţiuni anterioare a existat o presupoziţie implicită suplimentară, căreia i s-a permis să treacă neobservată până acum. Necesitatea acestei presupoziţii este *suficientă* pentru a anula orice sugestie conform căreia am putea deduce regula modulului pătrat al procedurii **R** din evoluţia U — chiar şi

FAPP. În chiar folosirea unei matrice densitate am *presupus*, într-adevăr, că un amestec de probabilitate ponderat este descris corespunzător printr-un astfel de obiect. Însăşi adecvarea unei expresii de forma $|\alpha><\alpha|$, care este ea însăşi o formă a „unui lucru sau a conjugatului său complex" este legată cu presupoziţia existenţei regulii modulului pătrat. Regula pentru obţinerea probabilităţilor dintr-o matrice densitate combină în mod corect probabilităţi clasice şi cuantice doar pentru că regula modulului pătrat este cuprinsă în chiar noţiunea de matrice densitate.

Deşi este adevărat că procesul de evoluţie unitară **U** este inseparabil, matematic, de noţiunea de matrice densitate şi de produsul scalar în spaţiul Hilbert $<\alpha|\beta>$, el nu ne spune în nici un fel că *probabilităţile* sunt cele care trebuie calculate prin metoda modulului pătrat. Este din nou o problemă de coexistenţă între **R** şi **U**, mai degrabă decât o explicaţie a lui **R** prin **U**. Evoluţia **U** nu spune absolut nimic despre noţiunea de probabilitate. Este o *presupoziţie* adiţională faptul că probabilităţile cunatice pot fi calculate prin acest procedeu, indiferent cum vrea cineva să justifice consistenţa lui **R** şi **U**, chiar dacă este vorba despre interpretarea lumilor multiple, sau despre abordarea FAPP.

Deoarece majoritatea suportului experimental al teoriei cuantice vine din felul în care teoria ne spune că trebuie calculate probabilităţile, nu putem ignora partea **R** a mecanicii cuantice. Este ceva diferit de **U**, este ceva care trebuie înţeles ca un proces fizic în sine. Cu asta nu vreau să spun că ar trebui să aibă o *lege* fizică specială. Fără îndoială, **R** este aproximarea a altceva, încă necunoscut. Discuţiile de la sfârşitul secţiunii precedente sugerează că folosirea procedurii **R** în procesul măsurării este într-adevăr o aproximaţie.

Să acceptăm deci că avem nevoie de ceva nou şi să încercăm să ne aventurăm mai departe, cu precauţia corespunzătoare, pe cariatele drumuri ale necunoscutului care se pot deschide în acest fel.

6.8. Este conştiinţa cea care reduce vectorul de stare?

Printre cei care iau $|\psi>$ în serios ca pe o descriere a lumii fizice, sunt unii care ar argumenta, ca o alternativă la folosirea lui **U** la toate nivelele şi la punctul de vedere al lumilor multiple, că ceva de natura lui **R** are loc în mod actual

doar dacă conştiinţa unui observator este implicată în procesul de măsurare. Distinsul fizician Eugene Wigner a creionat odată o teorie de acest gen (Wigner, 1961). Ideea generală ar fi că evoluţia materiei inconştiente — sau doar nevii - ar fi mereu guvernată de **U,** în momentul în care o entitate conştientă (sau o 'viaţă') ar fi legată fizic de sistem (neseparabilă), ar interveni ceva cu totul nou, un proces fizic al cărui rezultat este reducţia actuală a stării.

Nu este nevoie să mergem cu imaginaţia atât de departe încât să ne imaginăm că entitatea conştientă trebuie să fie în stare să 'influenţeze' alegerea particulară pe care o face Natura. O astfel de sugestie ne-ar duce în ape foarte tulburi şi, în ce mă priveşte, cred că există un conflict sever între faptele observate şi sugestia mult prea simplistă conform căreia un act de voinţă conştient poate influenţa rezulatul unui experiment al mecanicii cuantice. Astfel, *nu* cerem aici ca 'liberul arbitru conştient' să aibă vreun rol în privinţa lui **R** (vezi paragraful 7.1 pentru unele puncte de vedere alternative).

Fără îndoială că anumiţi cititori s-ar fi aşteptat ca, din moment ce caut o legătură între problema măsurătorii cuantice şi problema conştiinţei, să ader la idei de acest tip. Nu este cazul. Până una alta, conştiinţa este un fenomen rar în Univers. O bună parte pare să aibă loc la suprafaţa pământului, dar, în măsura în care avem deocamdată date experimentale [8], nu există în universul înconjurător, la distanţe de mulţi ani lumină de noi. Stranie ar mai fi imaginea unui univers fizic 'real' în care obiectele fizice ar evolua într-un mod cu totul diferit în funcţie de modul în care sunt sau nu în raza de percepţie a unuia dintre locuitorii săi conştienţi.

Să considerăm, de exemplu, vremea. Tipurile detaliate ale vremii care se dezvoltă pe orice planetă sunt dependente de procese haotice, deci sensibile la numeroase evenimente cuantice individuale. Dacă procesul **R** n-ar avea loc în absenţa conştiinţei, nici un comportament actual n-ar putea rezulta din infinitatea de superpoziţii distincte — ar fi o bulibăşeală totală, foarte diferită de ceea ce suntem obişnuiţi să numim timpul probabil — până când o fiinţă înzestrată cu conştiinţă devine conştientă de fenomen; şi *doar* în acest moment, superpoziţiile meteorologice devin în mod actual vreme. S-ar putea spune că din punct de vedere operaţional — din punctul de vedere operaţional al unei fiinţe conştiente, adică — o astfel de 'vreme' suprapusă n-ar fi câtuşi de puţin diferită de o vreme actuală nedeterminată (FAPP!). Totuşi, acest lucru nu este în sine o soluţie mulţumitoare

la problema realităţii fizice.

Astfel, propun să căutăm în altă parte pentru rezolvarea problemelor mecanicii cuantice. Cu toate că poate fi foarte bine cazul ca problema minţii să fie, în cele din urmă, legată de măsurătoarea cuantică – sau cu paradoxul **U**/**R** al mecanicii cuantice – cred că nu conştiinţa însăşi (sau cel puţin cea cu care suntem familiari sub acest nume) este cea care poate rezolva chestiunile fizice interne ale teoriei cuantice. Cred că problema măsurătorii cuantice trebuie înfruntată şi rezolvată înainte de a putea aştepta să facem vreun progres real în privinţa problemelor conştiinţei. O dată ce vom fi în posesia unei soluţii satisfăcătoare a problemei măsurătorii, vom fi într-o poziţie mai bună pentru a căuta un anume tip de răspuns privind chestiunea conştiinţei. Convingerea mea este că problema măsurătorii este o cerinţă preliminară pentru o înţelegere a minţii. Ele nu se reduc câtuşi de puţin una la alta. Problema minţii este mult mai dificilă decât problema măsurătorii!

6.9. Luând $|\psi\rangle$ în serios

Până acum, punctele de vedere care pretind că iau descrierea cuantică a realităţii în serios, nu reuşesc, după câte văd, s-o ia destul de în serios. Formalismul cuantic este poate prea straniu pentru a fi uşor să luăm în serios ceea ce spune şi cei mai mulţi dintre fizicieni se vor ruşina să interpreteze prea mult ce se întâmplă 'în realitate' în jurul lui. Pentru că pe lângă starea vectoare $|\psi\rangle$ care evoluează conform cu **U**, atâta timp cât sistemul rămâne la nivel cuantic, avem acţiunea perturbatoare şi probabilistă a lui **R**, care pare să producă 'sărituri' discontinue ale lui $|\psi\rangle$ imediat ce efectele cuantice sunt destul de mărite pentru a influenţa lucrurile la nivel clasic. Astfel, dacă e să considerăm că $|\psi\rangle$ furnizează o imagine a *realităţii*, trebuie să considerăm că aceste *salturi* sunt de asemenea reale, indiferent cât inconfort ne-ar produce un asemenea punct de vedere. Totuşi, dacă cineva este chiar atât de serios în privinţa realităţii descrierii cu ajutorul vectorului de stare, trebuie să fie pregătit să introducă o anumită schimbare (preferabil foarte subtilă) în regulile actuale ale teoriei cuantice. Pentru că operaţia **U** este incompatibilă cu **R** şi este nevoie cel puţin de o grămadă de muncă pentru a acoperi crăpăturile dintre descrierile la nivel cuantic şi la nivel

clasic.

De fapt, de-a lungul anilor, au existat câteva încercări neconvenţionale de a construi teorii coerente de-a lungul acestei linii de gândire. Şcoala maghiară, condusă de Karolyhazy la Budapesta a elaborat, din 1966, un punct de vedere în care efectele gravitaţionale duceau la un fel de procedură **R** ca la un fenomen fizic real (vezi, de asemenea, Komar, 1969). Urmând o linie oarecum diferită, Philip Pearle de la Hamilton College, Clinton, NY, USA a sugerat, în 1976, o teorie non-gravitaţională în care **R** are loc ca un fenomen fizic real. Mai recent, în 1986, o abordare interesantă a fost propusă de Giancarlo Ghirardi, Alberto Rimini şi Tulio Weber şi, în urma unor încurajări din partea lui John Bell, au fost dezvoltate numeroase alte sugestii şi îmbunătăţiri [9].

Înainte de a prezenta, în secţiunile care urmează, propriile mele preferinţe (şi care împrumută mult din schema Ghirardi-Rimini-Weber (GRW)), voi prezenta mai întâi această abordare. Ideea fundamentală este acceptarea realităţii lui $|\psi>$ şi exactitatea procedurilor U standard în cea mai mare parte a timpului. Atunci, funcţia de undă a unei particule individuale, iniţial localizate, libere, va tinde, conform ecuaţiei lui Schrödinger, să se împrăştie înainte în toate direcţiile spaţiului pe măsură ce timpul progresează (vezi figura 6.1) (amintiţi-vă că funcţia de undă a unei particule descrie factorii de pondere complecşi pentru diferite locaţii pe care particula le-ar putea avea; putem considera că graficele din figura 6.1 reprezintă schematic partea reală a acestui factor de pondere). Astfel, pe măsură ce timpul trece, particula devine mai puţin localizată. Noua trăsătură a schemei GRW este să presupună că există o foarte mică probabilitate ca, deodată, funcţia de undă să se multiplice de o funcţie cu vârf foarte ascuţit numită funcţie gaussiană cu un anumit factor de împrăştiere definit de parametrul σ Acest lucru este ilustrat în figura 6.2. Funcţia de undă a particulei devine instantaneu foarte localizată, gata să se împrăştie din nou. Probabilitatea ca vârful acestei gaussiene să fie într-un loc sau într-altul va fi proporţională cu modulul pătrat al valorii funcţiei de undă în acel loc. În acest fel, schema capătă consistenţă cu *regula modulului pătrat* a teoriei cuantice.

Cât de des se aplică această procedură? Aproximativ o dată la fiecare sută de milioane de ani! Să numim acest timp, perioada T. Atunci, într-un interval de o secundă, şansa ca reducţia stării să aibă loc va fi mai mică decât 10^{-15} (deoarece sunt aproximativ 3×10^{7} secunde într-un an). Astfel, pentru o singură particulă,

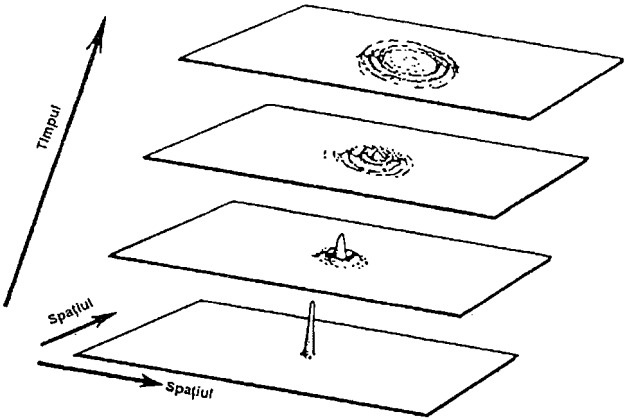


Fig. 6.1. Evoluţia temporală U a funcţiei de undă, iniţial localizată într-un punct şi care se răspândeşte apoi în toate direcţiile.

nu e un fenomen care să poată fi observat. Dar să ne imaginăm acum că avem un obiect destul de mare, ale cărui particule sunt fiecare obiectul unui asemenea proces. Pentru 10^{25} particule (cam atâtea câte sunt într-un şoarece mic), atunci şansa ca *anumite* particule să sufere un proces de acest tip va creşte enorm, perioada reducându-se la aproximativ 10^{-10} dintr-o secundă. Fiecare astfel de 'modificare' va afecta starea întregului sistem, deoarece fiecare particulă este non-separabilă de restul sistemului.

Să aplicăm acum această idee *pisicii lui Schrödinger* [10]. În acest paradox, -misterul X esenţial al teoriei cuantice — este vorba despre un obiect macroscopic (o pisică) plasat într-o superpoziţie liniară a două stări manifest diferite, ca pisică moartă sau ca pisică vie (vezi paragrafele 5.1 şi 6.6). Acest lucru este foarte uşor de făcut din perspective mecanicii cuantice, însă situaţia rezultantă pare de necrezut ca trăsătură a lumii *actuale* în care trăim — după cum Schrödinger nu a uitat să sublinieze (deşi unii adepţi realişti ai vectorului de stare ar lua mai degrabă drumul interpretării lumilor multiple sau al reducţiei 'conştiente' a vectorului de undă, aşa cum au fost ele descrise în paragrafele 6.2 şi 6.8). Pentru a construi o pisică a lui Schrödinger, nu avem nevoie decât de un eveniment cuantic de tip potrivit care să provoace o schimbare la nivel macroscopic — de fapt, o măsurătoare.

De exemplu, am putea avea un singur foton, emis de o sursă şi reflectat/transmis pe o oglindă semiargintată (ca în paragraful 5.7). Să spunem că partea transmisă a funcţiei de undă a fotonului declanşează un detector, cuplat de un dispozitiv care omoară pisisca, pe când partea reflectată iese din sistem şi nu face nici un rău pisicii (vezi figura 6.3). Ca şi în discuţia cu detectorii de mai înainte (paragraful 6.6), va rezulta o stare non-localizată cu o parte descriind pisica moartă, iar cealaltă, pisica vie şi fotonul care iese din sistem. Ambele posibilităţi vor fi luate *împreună* în vectorul de stare, atâta timp cât nu are loc un proces de reducţie **R**. Acest mister, misterul 'măsurării' este într-adevăr misterul **X** central al teoriei cuantice.

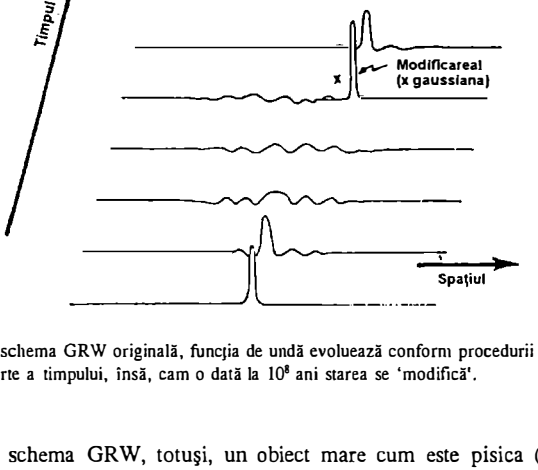


Fig. 6.2. În schema GRW originală, funcţia de undă evoluează conform procedurii U standard cea mai mare parte a timpului, însă, cam o dată la 10^8 ani starea se 'modifică'.

În schema GRW, totuşi, un obiect mare cum este pisica (care conţine aproximativ 10^{27} particule) trebuie să aibă aproape instantaneu una dintre particule modificată de o funcţie gaussiană (ca în figura 6.2) şi, cum starea acestei particule este neseparabilă de celelalte particule ale pisicii, reducţia uneia le va 'târâ' pe celelalte cu ea, determinând pisica să fie *ori* moartă *ori* vie. În acest fel, misterul **X** al pisicii lui Schrödinger este rezolvat, ca şi problema măsurătorii în general.

E o schemă ingenioasă, care suferă însă de pe urma caracterului ei ad-hoc.

Nu există nimic în celelalte părţi ale fizicii care să o susţină, iar valorile sugerate pentru T şi σ sunt pur şi simplu alese pentru a obţine rezultate 'rezonabile' (Diósi (1989) a sugerat o schemă asemănătoare, unde, parametrii T şi σ să fie fixaţi în funcţie de constanta gravitaţională a lui Newton, G ; între ideile sale şi cele pe care urmează să le descriu există o apropiere remarcabilă). O altă dificultate, mai serioasă, a explicaţiilor de acest tip este că produc o (mică) violare a principiului *conservării energiei*, lucru despre care voi mai vorbi în paragraful 6.12.

6.10. Reducţie a vectorului de stare indusă gravitaţional?

Există motive serioase[3] să presupunem că este nevoie de o modificare serioasă a teoriei cuantice, dacă este să considerăm **R** ca pe un proces fizic real şi, în plus, că această modificare trebuie să implice într-un mod serios efectele gravitaţiei. Unele dintre aceste motive au de-a face cu faptul că însuşi cadrul teoriei cuantice pare să fie în contradicţie cu noţiunile relativităţii generalizate, cum ar fi spaţiu-timpul curb.

Chiar concepte ca energia şi timpul, elemente de bază ale procedurilor fundamentale ale teoriei cuantice, nu pot fi definite în mod consistent cu cerinţele teoriei cuantice standard într-un context gravitaţional. Amintiţi-vă, de asemenea, efectul de 'înclinare' al conurilor luminoase (paragraful 4.4), specific fenomenului fizic al gravitaţiei. Ne-am putea aştepta, deci, la numite modificări ale principiilor de bază ale teoriei cuantice ca urmare a eventualei sale unificări cu relativitatea generalizată.

Şi totuşi, cei mai mulţi fizicieni au reţineri la ideea că teoria cuantică este cea care are nevoie de modificări pentru ca o astfel de uniune să se producă. Ei susţin că teoria lui Einstein este cea care trebuie modificată. Relativitatea generalizată clasică are propriile ei probleme, deoarece conduce la *singularităţi spaţio-temporale*, de tipul celor din găurile negre sau *Big-Bang*, unde curburile

[3] În MN, capitolele 7 şi 8, am prezentat astfel de motive în detaliu şi nu cred că este nevoie să repet argumentele aici. Ajunge dacă spun că toate acele argumente se pot susţine în continuare, chiar dacă criteriul particular pe care-l dau în paragraful 6.12 diferă întrucâtva de cel dat în MN.

spaţiu-timpului tind la infinit şi însăşi noţiunile de spaţiu şi timp îşi pierd validitatea (vezi MN, Capitolul 7).

Personal, nu mă îndoiesc că relativitatea generalizată va trebui modificată în interiorul unei teorii unificate şi aceste modificări vor fi esenţiale pentru a înţelege ce anume se întâmplă în acele regiuni pe care le descriem în prezent ca 'singularităţi' Acest lucru nu ne scuteşte însă de o modificare a teoriei cuantice.

Am văzut în paragraful 4.5 că relativitatea generalizată este o teorie foarte exactă nu mai puţin decât teoria cuantică însăşi. Cele mai multe dintre ingredientele fizice aflate la baza teoriei lui Einstein vor supravieţui cu siguranţă, într-o viitoare teorie unificată, aşa cum va supravieţui, probabil, cea mai mare parte a teoriei cuantice.

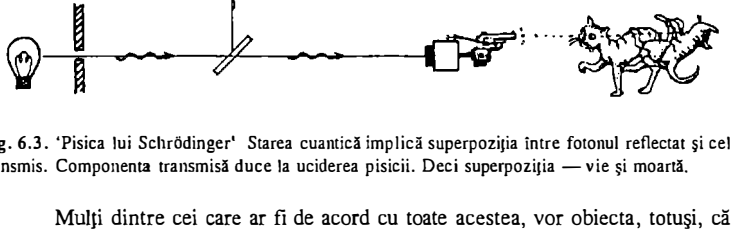


Fig. 6.3. 'Pisica lui Schrödinger' Starea cuantică implică superpoziţia între fotonul reflectat şi cel transmis. Componenta transmisă duce la uciderea pisicii. Deci superpoziţia — vie şi moartă.

Mulţi dintre cei care ar fi de acord cu toate acestea, vor obiecta, totuşi, că dimensiunile la care orice formă de gravitaţie cuantică ar fi relevantă sunt total nepotrivite cu ordinul de mărime ce caracterizează gravitaţia cuantică, numit *dimensiune Planck* (de 10^{-33} cm), de 20 de ori mai mică decât o particulă nucleară. Cum ar putea ceva ce se petrece la aceste dimensiuni să aibă ceva de-a face cu problema măsurătorii care se ocupă cu fenomene aflate la graniţa cu domeniul macroscopic. Cu toate acestea, avem de-a face cu o confuzie privind felul în care trebuie aplicate ideile gravitaţiei cuantice. Pentru că 10^{-33} este într-adevăr irelevant, dar nu în felul în care credeţi.

Să considerăm mai întâi un tip de situaţie cumva asemănător cu pisica lui Schrödinger, în care este produsă o stare în care o pereche de alternative posibil de distins macroscopic sunt liniar suprapuse. De exemplu, în figura 6.4 este imaginată o astfel de situaţie, unde un foton cade pe o oglindă semitransparentă,

starea fotonului devenind o superpoziţie a unei părţi transmise şi a unei părţi reflectate. Partea transmisă a funcţiei de undă activează (sau va activa) un dispozitiv care mişcă o bilă macroscopică dintr-o poziţie în alta. Câtă vreme evoluţia **U** se păstrează, 'locaţia' bilei se află într-o superpoziţie cuantică între ea însăşi în poziţia originală şi în poziţia deplasată. Dacă **R** este privit ca un proces fizic real, atunci bila va 'sări' fie într-o poziţie fie într-alta; şi acest proces constituie o 'măsurătoare' actuală. Ideea este că, în mod similar teoriei GRW, acest proces este fizic şi obiectiv şi va apărea ori de câte ori masa bilei sau distanţa pe care se deplasează sunt destul de mari. (În particular, acest proces nu are nimic de-a face cu intervenţia unei fiinţe conştiente care ar putea să 'perceapă' mişcarea sau bila.)

Îmi imaginez că aparatul care detectează fotonul şi mişcă bila este în sine destul de mic încât să poată fi în întregime tratat cu ajutorul mecanicii cuantice şi doar bila este cea care înregistrează măsurătoarea. De exemplu, în cazul limită, ne putem pur şi simplu imagina că bila este aşezată într-un echilibru instabil astfel încât simplul impact al fotonului va fi suficient pentru a-i provoca mişcarea.

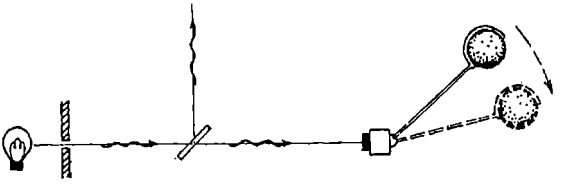


Fig. 6.4. În loc să avem o pisică, măsurătoarea poate consta dintr-o mişcare a unei lămpi sferice. Cât de masivă trebuie să fie lampa, sau cât de departe trebuie mutată?

Aplicând procedurile standard ale mecanicii cuantice (**U**), găsim că starea fotonului, după ce întâlneşte oglinda, constă din două părţi în două locaţii foarte diferite. Una dintre acestea devine non-separabilă cu aparatul şi, în final, cu bila, astfel încât avem o stare cuantică care implică o superpoziţie liniară a două poziţii foarte diferite ale bilei. Aceasta are propriul său câmp gravitaţional, care trebuie de asemenea considerat în superpoziţie. Conform teoriei lui Einstein, acest lucru implică faptul că avem două geometrii spaţio-temporale suprapuse! Întrebarea este dacă există un punct la care cele două geometrii devin suficient de diferite una de

alta încât regulile mecanicii cuantice să trebuiască să se schimbe şi, în loc de a obliga cele două geometrii diferite să intre în superpoziţie, Natura să aleagă între una şi alta şi să producă un anumit fel de reducţie asemănătoare lui **R**?

Problema este că nu avem nici o reprezentare a felului în care am putea concepe superpoziţiile liniare ale stărilor care presupun geometrii spaţio-temporale diferite. O dificultate fundamentală cu 'teoria cuantică' este că atunci când geometriile devin semnificativ diferite una de alta, nu avem absolut nici o metodă de a identifica un punct dintr-una cu vreun punct din cealaltă — cele două geometrii sunt spaţii strict *separate*- astfel că însăşi ideea de a forma o superpoziţie a unor stări *materiale* în aceste două spaţii separate devine profund obscură.

Ar trebui să ne întrebăm acum când trebuie cele două geometrii considerate 'semnificativ diferite' una de alta? Şi aici intervine dimensiunea Planck, de 10^{-33} cm. Ideea ar fi că diferenţa dintre cele două geometrii ar trebui să fie ceva de tipul 10^{-33} cm, sau mai mult (într-un sens adecvat) pentru ca reducţia să aibă loc. Am putea încerca să ne imaginăm, de exemplu (figura 6.5), că aceste două geometrii sunt forţate să coincidă, însă atunci când măsura diferenţei devine prea mare, de ordinul dimensiunii Planck, are loc reducţia **R** — cu alte cuvinte, în loc de a menţine superpoziţia, Natura trebuie să aleagă între o geometrie şi alta.

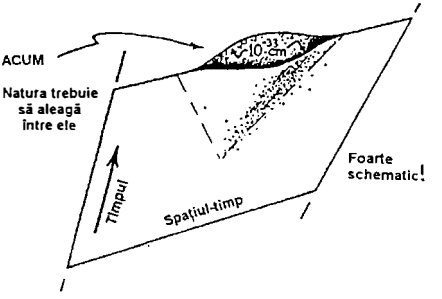


Fig. 6.5. Care este relevanţa dimensiunii Planck de 10^{-33} cm pentru reducţia vectorului de undă?

La ce fel de dimensiuni ale masei sau distanţei de deplasare corespunde această mică schimbare de geometrie? De fapt, datorită micimii efectelor gravitaţionale, acestea trebuie să fie destul de mari, deci posibil de conceput ca o linie de demarcaţie între nivelul clasic şi nivelul cuantic. Dar pentru a avea o idee

mai bine conturată despre toate acestea, ar fi util să spunem câte ceva despre *unităţile absolute (unităţile Planck)*.

6.11. Unităţi absolute

Ideea (datorată iniţial[4] lui Max Planck (1906) şi urmată în particular de John A. Wheeler (1975)) este de a folosi trei dintre constantele fizice fundamentale ale Naturii: viteza luminii c, constanta lui Planck (împărţită la 2π), $\hbar$ şi constanta gravitaţională a lui Newton, G, ca unităţi pentru a converti toate mărimile fizice în numere adimensionale. Asta revine la a alege unităţi de masă, lungime şi timp astfel încât toate aceste trei constante să devină unu: $c = 1$, $\hbar = 1$, $G = 1$

Dimensiunea Planck de 10^{-33} care, în unităţi de măsură obişnuite se exprimă ca $(G\hbar/c^3)^{1/2}$, ia acum valoarea 1, astfel încât este valoarea absolută de *lungime*. Unitatea absolută de *timp* corespunzătoare, timpul în care lumina traversează o distanţă Planck, este *timpul Planck* ($= (G\hbar/c^5)^{1/2}$), de aproximativ 10^{-43} secunde. Există de asemenea o unitate absolută de *masă*, numită *masă Planck* ($= (\hbar c/G)^{1/2}$), de aproximativ 2×10^{-5} grame, o masă foarte mare din punctul de vedere al fenomenelor cuantice obişnuite, dar destul de mică din perspectiva noastră obişnuită — ceva de ordinul masei unei muşte.

Evident, aceste unităţi nu sunt practice — poate cu excepţia masei Planck — însă pot fi utile atunci când considerăm efecte care pot avea relevanţă pentru gravitaţia cuantică. Iată câteva cantităţi relevante în unităţi absolute:

secunda $= 1{,}9 \times 10^{43}$
ziua $= 1{,}6 \times 10^{48}$
anul $= 5{,}9 \times 10^{50}$
metrul $= 6{,}3 \times 10^{34}$
cm $= 6{,}3 \times 10^{32}$
micronul $= 6{,}3 \times 10^{28}$

[4] O idee foarte asemănătoare a fost propusă cu 25 de ani mai târziu de fizicianul irlandez George Johnstone Stoney (1881), cu sarcina electronului luată ca unitate fundamentală (în locul constantei Planck, necunoscută la vremea aceea). Îi sunt recunoscător lui John Barrow pentru că mi-a indicat această sursă.

fermi (mărimea interacţiunilor puternice) = $6,3 \times 10^{19}$
masa nucleonului = $7,8 \times 10^{-20}$
gramul = $4,7 \times 10^{4}$
erg = $5,2 \times 10^{-17}$
gradul Kelvin = 4×10^{-33}
densitatea apei = $1,9 \times 10^{-94}$

6.12. Noul criteriu

Voi da acum un nou criteriu [11] pentru reducţia indusă gravitaţional a vectorului de stare, un criteriu care diferă semnificativ de cel sugerat în **MN** şi care este mai aproape de unele idei recente sugerate de Diósi şi alţii. *Motivaţia* pentru o legătură între gravitaţie şi procedura **R** dată în **MN** se păstrează încă, dar sugestia pe care o voi face se bucură de sprijin şi din alte direcţii. Ea este în plus eliberată de anumite probleme conceptuale cuprinse în definiţia de acolo şi este mult mai uşor de folosit. Propunerea din **MN** a fost de a găsi un criteriu după care cele două stări să poată fi considerate destul de diferite una de alta pentru a nu mai putea coexista într-o superpoziţie cuantică liniară (un criteriu cu privire la câmpurile lor gravitaţionale, adică la spaţiu-timpul fiecăreia). **R** ar fi trebuit să aibă loc în acest stadiu. Concepţia mea din acest moment este puţin diferită. Nu mai trebuie să căutăm o măsură absolută a diferenţei gravitaţionale între stări care să determine când ele sunt prea diferite pentru a putea coexista. În loc de asta, privim stările diferite suprapuse ca *instabile* — ca un nucleu de uraniu, de pildă — şi ne întrebăm care este *viteza* cu care această diferenţă determină reducţia vectorului de stare. Cu cât diferenţa este mai mare, cu atât mai rapidă va fi rata cu care reducţia are loc.

Pentru claritate, voi aplica noul criteriu în situaţia particulară care a fost descrisă în paragraful 6.10, însă poate fi uşor generalizat pentru a acoperi multe alte exemple. Ideea este de a vedea care este *energia* necesară pentru a separa una dintre stările bilei din exemplul de mai sus de cealaltă prin luarea în considerarea doar a efectului *gravitaţional*. Astfel, ne vom imagina că avem două bile (mase) care iniţial coincid şi se interpenetrează una pe alta (figura 6.6), după care vom

considera că una dintre stări se îndepărtează de alta, gradul de interpenetrare se reduce treptat, până când cele două ajung la starea de separaţie. Considerând inversul energiei gravitaţionale necesare pentru această operaţie, măsurat în unităţi absolute[5] obţinem timpul aproximativ necesar pentru apariţia reducţiei, de asemenea măsurat în unităţi absolute, timpul după care starea suprapusă a bilelor va trece spontan într-o stare localizată sau alta.

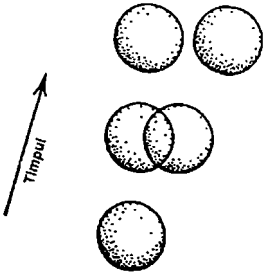


Fig. 6.6. Pentru a calcula timpul de reducţie $\hbar/E$, imaginaţi-vă că îndepărtăm una din ipostazele bilei de cealaltă şi calculăm energia din atracţia lor gravitaţională.

Dacă considerăm că bila este sferică, de masă m şi rază a, obţinem m^2/a pentru această energie. De fapt, valoarea actuală a energiei depinde de cât de departe trebuie să se mişte bila, însă această distanţă nu este foarte semnificativă dacă cele două ipostaze nu se suprapun prea tare atunci când ajung în poziţia finală. Energia adiţională necesară pentru mişcarea bilei din poziţia de contact (chiar până la infinit) este de acelaşi ordin de mărime (de 5/7 ori mai mare) cu cea necesară pentru mişcarea de la stadiul de conincidenţă la cel de contact. Astfel, în ce priveşte ordinele de mărime, putem ignora contribuţiile datorate deplasării bilelor după separaţie (dacă ele chiar se separă). Timpul de reducţie, conform acestei scheme, este de ordinul a a/m^2, măsurat în unităţi absolute, sau, schematic,

[5] Putem exprima timpul de reducţie în unităţi mai uzuale decât unităţile absolute adoptate aici. De fapt, expresia pentru timpul de reducţie este pur şi simplu $\hbar/E$, unde E este energia gravitaţională de separaţie la care ne-am referit înainte, fără ca vreo altă constantă în afara lui $\hbar$ să apară. Faptul că viteza luminii nu apare ne sugerează că un model 'newtonian' de această natură ar merita investigat (lucru făcut, de exemplu, de Christian, (1994)).

$$\frac{1}{10\varrho^2 a^5},$$

unde ϱ este densitatea bilei. Asta ne dă ceva de tipul $10^{186}/a^5$ pentru o densitate obişnuită (a unei picături de apă, să spunem).

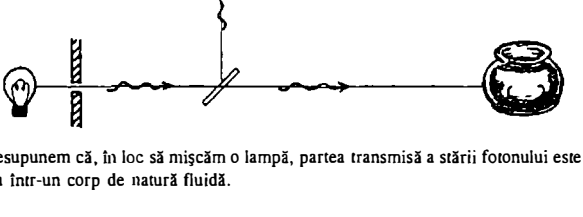


Fig. 6.7. Presupunem că, în loc să mişcăm o lampă, partea transmisă a stării fotonului este absorbită pur şi simplu într-un corp de natură fluidă.

Este liniştitor faptul că obţinem astfel nişte răspunsuri foarte 'rezonabile' în anumite situaţii extrem de simple. De exemplu, în cazul unui nucleon, (neutron sau proton) unde considerăm a 'mărimea interacţiunii puternice', 10^{-13} cm, adică, în unităţi absolute, 10^{20} şi luăm m de aproximativ, 10^{19}, obţinem un timp de reducţie de aproape 10^{58} ceea ce înseamnă peste zece milioane de ani. Este un timp încurajator de mare, pentru că au fost deja observate efecte de interferenţă pentru neutroni individuali [12]. Dacă am fi obţinut un timp de reducţie mai scurt, acest lucru ar fi fost în contradicţie cu astfel de observaţii.

Dacă considerăm ceva mai 'macroscopic', o mică picătură de apă de rază 10^{-5} cm, să spunem, timpul de reducţie va fi de ordinul orelor; pentru ceva şi mai mare, de ordinul unui micron, timpul ar fi a douăzecea parte dintr-o secundă; dacă raza obiectului este de 10^{-3} cm, timpul este mai puţin de a milioana parte dintr-o secundă. În general, atunci când considerăm un obiect aflat într-o superpoziţie de două stări spaţial distincte, trebuie doar să calculăm care este energia care ar fi produs deplasarea, în cazul în care între cele două nu există decât interacţiune gravitaţională. Inversul acestei energii măsoară ceva de tipul 'timpului de înjumătăţire' pentru starea suprapusă. Cu cât energia este mai mare, cu atât timpul cât starea suprapusă poate persista este mai mic.

Într-o situaţie experimentală propriu-zisă, ar fi foarte greu să păstrăm starea suprapusă a bilelor în izolare faţă de efectele perturbatoare ale mediului — caz în care am fi obligaţi să considerăm, de asemenea, efectele gravitaţionale implicate

în mediu. Acest factor suplimentar va fi relevant chiar dacă perturbarea nu produce mişcări macroscopice de masă în mediu; chiar micile deplasări ale particulelor individuale pot fi foarte importante (deşi, în mod normal, la o dimensiune totală mai mare a masei decât în cazul mişcării bilei macroscopice).

Pentru a clarifica efectul pe care o perturbaţie de acest tip îl poate avea în shema de faţă, să înlocuim bila din experimentul idealizat de mai înainte cu o masă fluidă care pur şi simplu *absoarbe* fotonul, dacă acesta este transmis prin oglindă (figura 6.7), astfel încât acum masa de fluid joacă rolul 'mediului' În loc să avem o superpoziţie liniară între două stări macroscopic distincte prin poziţiile lor spaţiale, avem acum două configuraţii de poziţii atomice diferite, în care o configuraţie de particule este deplasată aleator în raport cu cealaltă. Pentru o masă de material fluid oarecare de rază a, timpul de reducţie la care ne aşteptăm este de ordinul a $10^{130}/a^3$ (depinzând, până la un punct, de presupoziţiile făcute) în loc de $10^{186}/a^5$ cum obţinusem pentru mişcarea bilei. Aceasta sugerează că avem nevoie de mase ceva mai mari, pentru ca reducţia să aibă loc, decât în cazul unei bucăţi de materie care se mişcă. Cu toate acestea, însă, reducţia tot va apărea, chiar dacă nu există o mişcare macroscopică a întregii mase implicate.

Amintiţi-vă obstacolul material care intercepta raza fotonică din discuţia interferenţei cuantice în paragraful 5.8. Simpla *absorbţie* - sau chiar posibilitatea unei absorbţii — a unui foton de către un asemenea obstacol ar fi suficientă pentru a produce **R**, în ciuda faptului că nu se întâmplă nimic care să fie macroscopic observabil. Acest lucru ne arată că o perturbare suficientă într-un mediu *non-separabil* de sistemul considerat va produce **R**. Avem astfel o comunicare cu procedurile FAPP mai convenţionale.

Într-adevăr, în aproape toate procesele practice de măsurare este foarte probabil că numere mari de particule microscopice din mediul înconjurător vor fi perturbate. Conform ideilor propuse anterior, deseori *acesta* va fi efectul dominant şi nu mişcarea unor obiecte macroscopice. În absenţa unui mod de a controla situaţia experimentală foarte precis, orice mişcare macroscopică a unui obiect perceptibil va perturba o mare parte a mediului înconjurător şi este foarte probabil că timpul de reducţie al *mediului* (de aproximativ $10^{130}/b^3$, unde b este raza regiunii de densitatea apei a mediului non-localizat luat în consideraţie) va domina (adică va fi mult mai mic decât) timpul de care poate fi relevant pentru obiectul însuşi. De exemplu, dacă raza b a mediului perturbat ar fi de dimensiunea unei zecimi

dintr-un milimetru, reducţia ar avea loc într-un timp de ordinul unei milionimi de secundă.

O astfel de imagine are multe în comun cu descrierea convenţională din paragraful 6.6, dar acum avem un criteriu bine definit pentru ca **R** să apară în acest mediu. Amintiţi-vă obiecţiile care au fost discutate în paragraful 6.6 în imaginea FAPP convenţională de descriere a realităţii fizice actuale. Cu un astfel de criteriu obiecţiile acestea nu se mai susţin. Odată ce există o perturbare suficient de mare în mediu, conform ideilor prezente, reducţia va avea loc cu rapiditate în acel mediu, acompaniată de reducţia în orice aparat de măsură de care mediul este non-separabil. Nimic nu poate inversa reducţia şi să permită stării originale non-localizate să fie restaurată, chiar dacă ne-am putea imagina incredibile progrese ale tehnologiei. În consecinţă, nu există nici o contradicţie cu faptul că aparatul de măsură înregistrează *sau* DA *sau* NU.

Îmi imaginez că o descriere de această natură va fi relevantă în multe procese biologice şi că ea va furniza motivul pentru care structurile biologice de dimensiuni mult mai mici decât un micron în diametru pot să se comporte ca obiecte clasice. Un sistem biologic, fiind non-separabil de mediul său de maniera descrisă mai înainte, va avea propria sa stare redusă continuu, din cauza reducţiei continue a mediului. Pe de altă parte, ne putem imagina că din anumite motive, pentru sistemele biologice ar putea fi favorabil ca stările suprapuse să dureze pentru un timp mai lung. În astfel de cazuri ar fi necesar ca sistemul să fie, într-un fel, izolat efectiv de mediul înconjurător. Importanţa acestor consideraţii se va vedea mai târziu (paragraful 7.5).

Un punct care trebuie subliniat este că energia care defineşte timpul de viaţă al unei stări suprapuse este o *diferenţă* de energie şi nu energia totală implicată în situaţie. Astfel, pentru o bucată de materie destul de mare, dar nu se mişcă prea mult — dacă mai presupunem şi că, în plus, este cristalină, astfel încât atomii individuali nu sunt deplasaţi aleatoriu — atunci superpoziţiile cuantice pot fi menţinute pentru un timp destul de lung. Bucata de materie despre care este vorba poate fi mult mai mare decât picătura de apă considerată mai sus. Pot exista, de asemenea, mase mari de materie împrejur, cu condiţia să nu aibă loc non-localizări semnificative cu starea suprapusă care ne interesează (toate aceste consideraţii sunt importante pentru aparate cu corp solid ca, de exemplu, detectori de unde gravitaţionale care folosesc oscilaţiile coerente ale corpurilor solide

cristaline [13]).

Până aici, ordinele de mărime par destul de plauzibile, dar mai este desigur nevoie de multă muncă pentru a vedea dacă ideea va trece de o examinare ceva mai minuţioasă. Un test crucial ar fi să găsim situaţii experimentale în care teoria standard ar prezice efecte care depind de superpoziţii cuantice la distanţă, însă la un nivel la care abordarea anterioară susţine că superpoziţia nu poate fi menţinută. Dacă aşteptările convenţionale ale teoriei cuantice sunt susţinute de observaţii, atunci ideea mea va trebui abandonată — sau, cel puţin, modificată sever. Dacă observaţiile indică faptul că superpoziţiile nu sunt menţinute, atunci acest lucru ne furnizează sprijin pentru ideile de mai sus. Din păcate, nu ştiu să existe încă sugestii pentru asemenea experimente. Superconductorii şi aparatele de tip SQUID (care depind de superpoziţiile cuantice la scară mare care apar în superconductori) par să prezinte o arie de investigaţie promiţătoare (vezi Leggett, 1984). Totuşi, ideile pe care le susţin au nevoie de elaborare înainte de a putea fi direct aplicate unor situaţii de acest fel. În cazul superconductorilor o foarte mică cantitate de masă este deplasată între stările suprapuse. În loc de asta, avem o diferenţă semnificativă de impuls, ceea ce implică o dezvoltare teoretică a ideilor precedente pentru a acoperi şi această situaţie.

În plus, ideile de mai sus trebuie reformulate, chiar pentru a trata o situaţie simplă ca a camerei cu ceaţă — în care prezenţa unei particule încărcate este semnalată prin condensarea unor mici picături de apă din vaporii ambianţi. Să presupunem că avem o stare cuantică a unei particule încărcate constând dintr-o superpoziţie liniară a particulei într-un anumit punct al camerei şi a particulei în afara camerei. Partea din starea vectoare care exprimă prezenţa în interiorul camerei iniţiază formarea unei picături, dar nu şi partea care este în afara camerei, astfel încât starea constă într-o superpoziţie de două stări macroscopice diferite. Într-una, avem o picătură care se formează, pe când în cealaltă, în camera cu ceaţă rămân doar vapori uniformi. Trebuie să estimăm energia gravitaţională implicată în 'separarea' moleculelor de vapori între cele două stări iniţial suprapuse. Dar, avem în acest caz o complicaţie adiţională pentru că există o diferenţă între energia gravitaţională proprie a picăturii de apă şi vaporii necondensaţi. Pentru a rezolva astfel de situaţii, este utilă o formulare *diferită* a criteriului de mai înainte. Putem considera *energia gravitaţională proprie* a distribuţiei de masă, care este *diferenţa* dintre distribuţiile de masă ale celor două stări care sunt considerate în superpoziţie

liniară. Inversa acestei energii ne dă o propunere alternativă pentru dimensiunile timpului de reducţie (conform cu Penrose, 1994b). De fapt, această formulare alternativă ne dă exact acelaşi rezultat ca mai înainte, în situaţiile considerate anterior, doar că ne dă un timp de reducţie mai scurt în cazul camerei cu ceaţă. Într-adevăr, există diferite scheme alternative generale pentru timpul de reducţie, cu răspunsuri diferite în anumite situaţii, dar cu acelaşi tip de răspuns pentru exemplul cu bila considerat la începutul paragrafului. Schema originală îi aparţine lui Diósi (1989) (şi conţine câteva dificultăţi, după cum au arătat Ghirardi, Grassi şi Rimini (1990), care au sugerat şi un remediu). Nu voi discuta aceste propuneri aici, dar, în următoarele capitole, atunci când mă voi referi la 'propunerile din paragraful 6.12' mă voi referi şi la ele.

Care sunt însă motivaţiile pentru propunerea specifică pentru un 'timp de reducţie' despre care a fost vorba? Motivaţiile mele iniţiale (Penrose 1993a) erau prea tehnice pentru a le descrie aici şi erau, în plus, incomplete şi neconcluzive [14]. Voi prezenta un exemplu independent al acestui tip de schemă în continuare. Chiar dacă incompletă, argumentaţia pare să sugereze existenţa unor cerinţe de consistenţă subiacente foarte puternice care ne îndreptăţesc de asemenea să credem că reducţia de stare trebuie să fie, în ultimă instanţă, un fenomen gravitaţional de natura celei propusă aici.

Problema *conservării energiei* în schemele de tip GRW a fost deja amintită în paragraful 6.9. 'Modificările' la care funcţia de undă a particulei este supusă (când este multiplicată spontan cu o gaussiană) provoacă mici violări ale legii conservării energiei. Mai mult, pare să aibă loc un transfer non-local de energie în cadrul acestui proces. Pare să fie vorba de o trăsătură caracteristică (şi aparent imposibil de evitat) a teoriilor de acest tip, în care procedura **R** este considerată un fenomen fizic real. După părerea mea, asta înseamnă o evidenţă adiţională pentru teoriile în care efectele gravitaţionale joacă rolul cheie în procesul de reducţie. Pentru că principiul conservării energiei este o chestiune subtilă şi evanescentă şi în teoria relativităţii generalizate. Câmpul gravitaţional conţine el însuşi energie, care contribuie în mod măsurabil la energia totală a unui sistem (şi deci la masă, conform formulei lui Einstein, $E=mc^2$). Însă este vorba despre o formă nebuloasă de energie care există în spaţiul vid într-un mod misterios şi non-local. Amintiţi-vă că masa-energie este ceea ce e emis, sub formă de unde gravitaţionale, de către pulsarul binar PSR 1913 + 16 (vezi paragraful 4.5); aceste unde sunt deformări ale

structurii spaţio-temporale. Energia conţinută în câmpurile atractive a două stele neutronice este de asemenea un ingredient important şi care nu poate fi ignorat al dinamicii acestora. Însă acest fel de energie care există în spaţiul vid este în mod special problematică: nu poate fi obţinută prin 'adunarea' contribuţiilor locale ale densităţii de energie, nici nu poate fi localizată în vreo regiune particulară a spaţiu-timpului (vezi **MN**, p. 240). Este destul de tentant să legăm chestiunea problematică a energiei non-locale în procedura **R** de problemele gravitaţiei clasice şi să le folosim ca dovezi care se completează reciproc în favoarea unei imagini unificatoare coerente.

Dar ating sugestiile mele această coerenţă? Cred că au, cel puţin, şansa de a fi pe drumul cel bun. Evident, cadrul precis pentru obţinerea acestor rezultate nu există încă. Avem doar o bază pentru el. Aşa cum am văzut înainte, putem considera procesul de reducţie ca pe ceva asemănător cu dezintegrarea unui nucleu radioactiv. Imaginaţi-vă cele două stări ale bilei din exemplul nostru ca pe un nucleu instabil care se dezintegrează, după un 'timp de viaţă' caracteristic, în ceva mai stabil. Poziţiile suprapuse ale bilei poat fi considerate o stare cuantică care se dezintegrează într-o stare în care bila este într-o locaţie, sau o alta în care bila este în celălalt loc (cele două moduri de dezintegrare posibile).

În dezintegrarea particulelor sau nucleelor, timpul de viaţă (timpul de înjumătăţire) al procesului de dezintegrare este inversul unei mici indeterminări în energia particulei iniţiale — efectul principiului de nedeterminare al lui Heisenberg (de exemplu, masa unui nucleu de poloniu 210 instabil, care se dezintegrează cu emisia unei particule alfa, este definită cu o nedeterminare de ordinul inversului timpului de dezintegrare — în acest caz, de aproximativ 138 de zile, ceea ce ne dă o incertitudine de 10^{-34} a masei. Pentru particulele individuale instabile, nedeterminarea este mult mai mare.) Astfel, 'dezintegrarea' implicată în procesul de reducţie ar trebui de asemenea să implice o nedeterminare esenţială în energia stării iniţiale. Această nedeterminare, conform ideii de mai înainte, se datorează în mod esenţial în incertitudinea energiei gravitaţionale proprii a stării suprapuse. O astfel de energie gravitaţională proprie implică şi nebuloasa energie a câmpului non-local care provoacă atâtea dificultăţi în relativitatea generalizată şi care nu poate fi obţinută prin adunarea densităţilor locale de energie. Ea implică de asemenea nedeterminarea esenţială a identificării punctelor a două geometrii spaţio-temporale diferite la care ne-am referit în paragraful 6.10. Dacă privim într-

adevăr contribuţia gravitaţională ca factorul esenţial în energia stării suprapuse, rezultatul se potriveşte cu timpul de viaţă propus pentru acea stare. Astfel, schema prezentată pare să furnizeze o legătură clară între cele două probleme privind energiile şi acest lucru sugerează destul de convingător direcţia în care ar trebui căutată o teorie unificată.

În sfârşit, există două întrebări importante de relevanţă specială în acest moment. Prima este: ce rol pot avea aceste consideraţii pentru acţiunea *creierului*? A doua: există vreun motiv să ne aşteptăm, pe baze exclusiv fizice, ca non-computabilitatea să fie o trăsătură a acestui proces de reducţie indus gravitaţional? Vom vedea, în capitolul următor, că există, într-adevăr, perspective şi posibilităţi interesante.

Note şi referinţe bibliografice

1. Există un sens în care proprietatea 'bosonică' a fotonilor la care ne-am referit în 5.16, poate fi privită ca un exemplu de non-separabilitate cuantică -- caz în care observaţiile lui Hanbury Brown şi Twiss (1954,1956) oferă o confirmare pentru distanţe mari (vezi nota din paragraful 5.16)
2. Everett (1957), Wheeler (1957), DeWitt şi Graham (1973), Geroch (1984).
3. Squires (1990,1992b)
4. Bell (1992)
5. Pentru un argument diferit care susţine realitatea obiectivă a funcţiei de undă vezi Aharonov, Anandan şi Vaidman (1993).
6. Vezi, de exemplu, d'Espagnat (1989).
7. D'Espagnat (1989), Zurek (1991, 1993), Paz, Habib, Zurek (1993).
8. Aceasta pare să fie concluzia programului SETI condus de F. Drake.
9. Propriile mele sugestii, cu toate că situate în mod hotărât în tabăra 'gravitaţională', n-au fost prea exacte până de curând, cf. Penrose (1993a, 1994b). Această propunere împarte cu cea originală formulată de Ghirardi-Rimini-Weber idea că reducţia trebuie să fie un proces rapid şi discontinuu. O mare parte din activitatea curentă, totuşi, se ocupă cu procese de reducţie *continue* (stochastice), ca cele propuse de Pearle (1976). Vezi Diósi(1992), Ghirardi et.al (1990b), Percival (1994). Pentru lucrări de această natură care cer să facă schema consistentă cu relativitatea, vezi Ghirardi et.al, (1992), Gisin (1989), Gisin şi Percival (1993).
10. Schrödinger (1935a), vezi şi MN.

11. Diósi (1989), Ghirardi et.al., (1990a), Penrose (1993a).
12. Zeilinger et.al., (1998)
13. Weber (1960), Braginski (1977).
14. Totuşi, motivaţiile generale date în **MN**, Capitolul 7, par să sprijine propunerea de aici (vezi şi Penrose 1993a) cu mai multă claritate decât acestea sprijină criteriul *one-gravitation*, aşa cum a fost dat în **MN**. Sunt necesare cercetări suplimentare pentru a face legăturile mai clare.
15. Vezi Penrose (1991a); de asemenea **MN** p.240

7. TEORIA CUANTICĂ ŞI CREIERUL

7.1. Acţiune cuantică la scară mare în funcţionarea creierului?

Acţiunea creierului trebuie înţeleasă, în interpretarea convenţională, în termeni esenţialmente clasici — sau cel puţin aşa ar părea. Semnalele nervoase sunt în mod normal considerate fenomene on/off (pornire *sau* oprire), asemănător curenţilor electrici dintr-un calculator: ori există ori nu există, fără vreuna din acele misterioase alternative *suprapuse* caracteristice acţiunilor cuantice.

Până când vom admite că, la nivelele microscopice din creier, efectele cuantice trebuie să aibă rolul lor, biologii par să împărtăşească opinia generală după care nu este necesar să ne modificăm imaginea clasică atunci când avem de discutat implicaţiile la scară mare ale acestor ingredienţi cuantici primitivi.

Forţele chimice care controlează interacţiunile atomilor şi moleculelor sunt într-adevăr cuantice la origine; de asemenea, acţiunea chimică guvernează, în general, comportamentul substanţelor *neurotransmiţatoare* care transferă semnale de la un neuron la altul — de-a lungul unor mici goluri, numite *fante sinaptice*. Similar, potenţialele de acţiune care controlează transmisiunea semnalelor nervoase au, de asemenea, o origine cuantică. Totuşi, opinia generală pare să fie cea după care este destul de adecvată modelarea comportamentului neuronilor şi a relaţiilor dintre ei într-un mod complet clasic. În consecinţă, este extrem de răspândită convingerea conform căreia este posibilă modelarea funcţionării creierului ca un sistem *clasic*, în care trăsăturile subtile şi misterioase ale fizicii cuantice nu sunt

luate în consideraré.

Acest mod de abordare are ca implicaţie faptul că orice activitate semnificativă care ar putea avea loc în creier, nu poate fi decât de tipul apare *sau* nu apare. Superpoziţiile stranii ale teoriei cuantice care permit 'apariţia' *şi* 'neapariţia' simultană — cu factori de pondere numere complexe — nu vor fi luate în considerare, chiar dacă se acceptă faptul că la nivel submicroscopic ele se pot într-adevăr produce. Astfel se va considera adecvată tratarea oricăror superpoziţii de acest fel de parcă ar fi vorba despre amestecuri statistice, iar modelarea clasică a activităţii creierului va fi perfect satisfăcătoare FAPP.

Desigur, există şi unele puncte de vedere diferite. În particular, celebrul neurofiziolog, John Eccles a susţinut importanţa efectelor cuantice în acţiunea sinaptică (vezi, în particular, Beck şi Eccles (1992), Eccles, (1994)). El a indicat ca posibil 'loc' al acţiunii cuantice, reţeaua veziculară presinaptică o reţea paracristalină hexagonală în celulele hexagonale ale creierului. Alţii, (printre care şi eu, vezi Penrose, 1987 şi MN) au încercat să folosească faptul că celulele sensibile la lumină de pe retină (care, din punct de vedere tehnic este o parte a creierului) pot răspunde la un număr mic de fotoni (Hecht *et.al.* 1941) sau chiar la un *singur* foton (Baylor *et.al*, 1979), în condiţii speciale pentru a formula ipoteza conform căreia ar putea exista şi în creier neuroni care ar putea de asemnea să 'joace' rolul unor detectori cuantici.

Împreună cu posibilitatea ca efectele cuantice să fie implicate la nivele mult mai importante ale acţiunii creierului, unii au exprimat, de asemenea, speranţa ca în asemenea circumstanţe, *indeterminarea cuantică* să fie mecanismul care ar putea oferi o deschidere către influenţa *minţii* asupra creierului fizic. În acest context este probabilă adoptarea unui punct de vedere *dualist*, explicit sau implicit.

Poate că *liberul arbitru* unei 'minţi externe' ar putea fi în stare să influenţeze alegerile cuantice care rezultă din procesele nedeterministe. Din acest punct de vedere, 'substanţa mentală' [1] va influenţa comportamentul creierului

---

[1] Tot acest paragraf este o referire la teoriile care se inspiră din dualismul cartezian, după care *substanţa gânditoare* este ceva fundamental 'diferit' de materia din care este alcătuit creierul, dar îl poate 'comanda' pe acesta. Creierul este doar sediul fizic al unei *substanţe* nemateriale. În teoriile născute din această presupoziţie, problema esenţială este cum anume pot interacţiona cele două substanţe complet diferite. Penrose face referire la procesul de reducţie a vectorului de undă ca la unul dintre mecanismele posibile. (n.t.)

exact prin intermediul procesului R.

Statutul unor asemenea sugestii îmi este oarecum neclar, în special de când, în teoria cuantică standard, indeterminarea cuantică *nu* apare la nivel microscopic, unde se păstrează întotdeauna evoluţia deterministă **U**. Doar în procesul de amplificare de la nivel cuantic la nivel clasic apare nedeterminarea procesului **R**. Conform punctului de vedere standard FAPP, această indeterminare este ceva ce 'are loc' doar atunci când cantităţi suficiente din mediul înconjurător sunt 'legate' cuantic de evenimentul cuantic. De fapt, aşa cum am văzut în paragraful 6.6, în interpretarea standard a mecanicii cuantice nu este clar nici măcar ce înseamnă 'a avea loc'.

Pornind de la imaginea convenţională, ar fi dificil să susţinem că teoria permite cu adevărat o indeterminare la nivelul unei singure particule cuantice (foton, atom, moleculă). Atunci când, de exemplu, o funcţie de undă a unui foton întâlneşte o celulă fotoelectrică sensibilă declanşează o secvenţă de evenimente care rămâne esenţial deterministă (sub acţiunea lui **U**) atâta timp cât se poate considera că sistemul rămâne 'la nivel cuantic'. În cele din urmă, o parte seminficativă a mediului este perturbată şi, în interpretarea convenţională, considerăm că **R** a apărut FAPP. Va trebui deci să ne mulţumim cu faptul că 'substanţa metală' influenţează cumva sistemul doar la acest nivel nedeterminat.

Conform punctului de vedere asupra reducţiei vectorului de undă pe care l-am susţinut în această carte, (vezi paragraful 6.12), pentru a găsi nivelul la care procesul **R** devine operativ în mod *actual*, trebuie să ne uităm la dimensiuni mai mari, care devin relevante atunci când o cantitate considerabilă de material (dimensiuni de ordinul micronilor sau milimetrilor în diametru sau poate chiar mult mai mult, dacă nu este implicată o mişcare semnificativă a masei) este nonlocalizată în starea cuantică (voi denumi de aici înainte această procedură specifică prin **OR**[2], ceea ce înseamnă *reducţie obiectivă*)[3]. În orice caz, dacă

---

[2] În original, "objective reduction", în prescurtare **OR**, cu semnificaţia "sau" (ori..ori). Un joc de cuvinte pe care nu l-am putut păstra (n.t.)

[3] În MN am folosit eticheta "gravitaţie cuantică corectă" (GCC) pentru acest gen de lucru. Aici accentul este pus oarecum diferit. Nu vreau să subliniez legătura dintre acest procedeu şi nevoia de a găsi o teorie complet coerentă a gravitaţiei cuantice. În acest moment vreau să subliniez o procedură de felul celei dezvoltate cu ajutorul sugestiilor formulate în paragraful 6.12, însă având în plus un ingredient non-computaţional fundamental. Folosirea noii denumiri **OR** are calitatea adiţională că ne

încercăm să aderăm la punctul de vedere dualist expus înainte, unde căutăm un mecanism prin care o 'minte externă' să poată influența comportamentul fizic al creierului — probabil prin înlocuirea nedeterminării pure a teoriei cuantice prin ceva mai subtil atunci trebuie într-adevăr să găsim modul în care influența 'minții' poate avea loc la o scară mult mai mare decât particulele cuantice individuale. Trebuie să căutăm în locurile în care apar puncte de intersecție între nivelul clasic și nivelul cuantic. Așa cum am văzut în capitolul precedent, nu există un acord asupra a ceea ce este un asemenea punct de intersecție, unde este el, sau dacă există.

După părerea mea, nu este foarte util din punct de vedere științific să ne gândim la o 'minte' dualistă care ar fi (logic) *externă* corpului, influențând cumva posibilitățile care par să apară în acțiunea lui R. Dacă 'voința' poate influența cumva felul în care Natura alege dintre alternaltivele care apar din R, atunci de ce nu ar fi în stare experimentatorul, prin 'puterea voinței', să influențeze rezultatul unui experiment cuantic? Iar dacă acest lucru ar fi posibil, atunci ar trebui să fie evidente violările probabilităților cuantice. În ce mă privește, nu pot crede că o asemenea perspectivă poate fi adevărată. Postularea unei 'substanțe mentale' externe, care nu este ea însăși supusă legilor fizice, ne scoate în afara oricărei explicații științifice posibile și înseamnă aderarea la punctul de vedere 3 (vezi paragraful 1.3).

Este totuși greu să argumentăm împotriva acestui punct de vedere într-un mod riguros, deoarece el este, prin însăși natura sa, eliberat de regulile clare care i-ar permite încadrarea printre subiectele unei dispute științifice. Acelor cititori care sunt convinși, indiferent de motiv, că știința va fi mereu incapabilă să trateze chestiuni legate de minte, le cer totuși să rămână alături de mine pentru a vedea dacă nu cumva am putea găsi un loc pentru această descriere în interiorul unei științe extinse mult dincolo de scopul limitat pe care îl are astăzi. Dacă 'mințe' este ceva extern corpului fizic, este greu de văzut de ce atât de multe din atributele sale pot fi strâns asociate proprietăților creierului. Părerea mea este că ar trebui să explorăm ceva mai adânc în structurile fizice actulale, 'materiale', care constituie creierul și, de asemenea, să reconsiderăm întrebarea „Ce anume este o structură

---

aduce aminte de faptul că într–o procedură de reducție obiectivă, rezultatul fizic este, într–adevăr, fie un lucru, fie altul, în locul unei superpoziții combinate.

'materială' la nivelul cuantic al lucrurilor?" Cred că în final nu există scăpare în faţa necesităţii de a reexamina adevărurile care se află ascunse la rădăcinile Naturii.

Oricum am lua-o, cel puţin un lucru e sigur. Nu trebuie să privim doar spre efectele cuantice ale particulelor individuale – atomi, molecule de dimensiune redusă- ci spre efectele sistemelor cuantice care îşi păstrează natura cuantică manifestă la o scară mult mai largă. Dacă nu există o coerenţă cuantică la scară mare, atunci e imposibil ca vreunul dintre subtilele efecte de la nivel cuantic – non-localitate, paralelism cuantic (faptul că mai multe acţiuni susprapuse au loc simultan) efecte ale contrafactualităţii să aibă vreo relevanţă în activitatea creierului la nivel clasic. Fără o 'ecranare' adecvată a stării cuantice, astfel de efecte vor fi imediat pierdute în comportamentul inerent aleator al mediului adică, în mişcările aleatoare ale acelor materiale şi fluide biologice care constituie cea mai mare parte a creierului.

Ce este *coerenţa cuantică*? Acest fenomen se referă la împrejurările în care un număr mare de particule pot coopera colectiv într-o singură stare cuantică care rămâne separată de influenţa mediului (cuvântul coerenţă se referă, în general, la faptul că oscilaţiile în diferite locuri sunt în fază; aici, în cazul coerenţei *cuantice* ne interesează natura oscilatorie a funcţiei de undă, iar coerenţa se referă la faptul că lucrăm cu o singură stare cuantică). Astfel de stări apar cel mai dramatic în fenomenul de superconductivitate (unde rezistenţa electrică scade la zero) şi superfluiditate (vâscozitatea scade la zero). Ingredientul caracteristic al unui asemenea fenomen este prezenţa unui *gol energetic* (bandă de energie interzisă) care trebuie 'sărit' de particulele din mediu pentru ca starea cuantică să poată fi perturbată. Dacă temperatura mediului este prea mare, astfel încât energia multora dintre particule este destul de mare pentru a 'sări' golul de energie şi a interacţiona cu starea cuantică, coerenţa cuantică este distrusă. În consecinţă, fenomene de felul superconductivităţii sau superfluidităţii au fost puse în evidenţă în mod normal numai la temperaturi foarte joase, de câteva grade peste zero absolut. Din aceste motive exista până nu demult un scepticism generalizat privind posibilitatea relevanţei efectelor de coerenţă cuantică pentru nişte obiecte atât de 'fierbinţi' ca creierul uman, sau oricare alt sistem biologic.

În ultimii ani, însă, anumite rezultate experimentale remarcabile au arătat că, pentru anumite substanţe, superconductivitatea poate apărea la temperaturi mult mai

mari, până la 115K (conform Sheng *et.al.* 1988). Această temperatură este încă foarte mică din punct de vedere biologic, (-158 °C, doar ceva mai cald decât azotul lichid). Dar şi mai interesante sunt observaţiile lui Langues *et.al.* (1993) care par să indice prezenţa superconductivităţii la temperaturi 'siberiene' de -23°C. Deşi încă într-un domeniu cam rece pentru biologie, această *superconductivitate la temperaturi ridicate* ne dă un important sprijin în favoarea ipotezei că pot exista efecte de coerenţă cuantică care să fie cu adevărat relevante pentru sistemele biologice.

De fapt, cu mult înainte ca fenomenul de superconductivitate la temperatură înaltă să fie observat, fizicianul Herbert Fröhlich (cel care, în anii '30 a realizat unul dintre paşii importanţi în înţelegerea superconductivităţii 'obişnuite' la temperaturi scăzute) a sugerat posibilitatea ca efectele cuantice colective să joace un rol în sistemele biologice. Cercetarea sa a fost stimulată de un fenomen enigmatic observat deja în 1938, în membranele biologice. Astfel, în 1968 Fröhlich a propus (folosind un concept datorat fratelui meu Oliver Penrose şi lui Lars Onsager, (1956) cum am aflat şi eu din întâmplare uitându-mă la aceste lucrări) posibilitatea unor efecte vibraţionale în interiorul celulelor active, efecte care ar rezona cu radiaţia electromagnetică de microunde, la 10^{11} Hz, ca rezultat al fenomenului de coerenţă cuantică biologic. În loc să avem nevoie de o temperatură scăzută, efectele apar din existenţa unei energii mari a proceselor metabolice. Există astăzi anumite evidenţe observaţionale demne de luat în considerare pentru exact acel tip de efect pe care Fröhlich l-a prezis în 1968. Vom vedea mai târziu (în paragraful 7.5) ce relevanţă poate avea acest fenomen asupra acţiunii creierului.

7.2. Neuroni, sinapse şi calculatoare

Până acum, cu toate că este încurajator faptul că am găsit o posibilitate distinctă ca fenomenele de coerenţă cuantică să joace un rol semnificativ în sistemele biologice, nu există o legătură directă între acest lucru şi ceea ce poate fi direct relevant pentru activitatea creierului.

O mare parte din înţelegerea creierului, încă destul de rudimentară, ne-a condus la imaginea clasică (în mare parte aceeaşi cu cea propusă în 1943

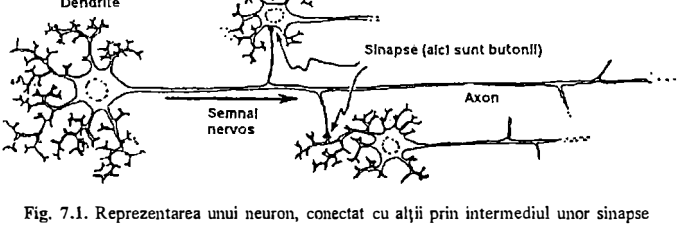


Fig. 7.1. Reprezentarea unui neuron, conectat cu alţii prin intermediul unor sinapse

de McCullogh şi Pitts) în care neuronii şi sinapsele par să joace un rol esenţial, similar celui al tranzistorilor şi cablajelor de legătură în calculatoarele de astăzi. Mai detaliat, descrierea biologică este a unor semnale nervoase care se propagă pornind dinspre capul neuronului de-a lungul unei fibre extrem de lungi numite *axon*, care se bifurcă în ramuri separate în diferite locuri. Fiecare ramură se termină la o *sinapsă*, joncţiunea la care semnalul este transferat, de obicei unui alt neuron, peste o fantă sinaptică. În acest stadiu, substanţele chimice neurotransmiţătoare duc mesajul pe care l-a emis primul neuron, trecând de la o celulă (neuron) la alta. Joncţiunea sinaptică apare adesea la dendrita (în formă arborescentă) celuilalt neuron, sau în corpul (capul) său. Anumite sinapse sunt de natură excitatoare, cu neorotrasmiţători care tind să mărească probabilitatea ca al doilea neuron să emită, pe când altele sunt inhibitoare, iar diferitele lor substanţe neurotransmiţătoare tind să inhibe capacitatea celui de-al doilea neuron de a emite. Efectele diferitelor acţiuni sinaptice asupra neuronului următor se adună ('plus' pentru excitaţie, 'minus' pentru inhibiţie), iar când se atinge un anumit nivel, următorul neuron va emite[4]. Mai corect ar fi, de fapt, să spunem că există o mare posibilitate ca acesta să emită. În toate aceste procese sunt în mod obligatoriu implicaţi şi nişte factori aleatori.

Până acum nu este nici o îndoială că această imagine este una ce poate fi

[4] Cel puţin aceasta a fost imaginea convenţională. Există acum dovezi că simpla descriere 'aditivă' poate fi o supra–simplificare considerabilă şi că este foarte probabil ca la nivelul dendritelor neuronilor individuali să existe o anumită 'procesare a informaţiei'. Această posibilitate a fost subliniată de Karl Pribram şi alţii (vezi Pribram, 1991). Anumite sugestii preliminare în sensul acesta au fost făcute de Alwyn Scott (1973, 1977, iar pentru posibilitatea *inteligenţei* în interiorul celulelor individuale vezi Albrecht Buefler, 1985, de exemplu). Că poate să existe o „procesare dendritică complexă" care are loc în neuronii individuali – acesta este un lucru consistent cu discuţia din paragraful 7.4.

în principiu simulată computaţional, presupunând fixate conexiunile sinaptice şi 'tăriile' lor individuale (elementele aleatoare nu vor pune probleme computaţionale, desigur, conform paragrafului 1.9). Într-adevăr, nu este greu de văzut că descrierea pe bază de sinapse neuronice prezentată aici (cu sinapse fixate) este în mod *esenţial* echivalentă cu cea a unui calculator (vezi **MN** p. 420-424). Însă, datorită unui fenomen cunoscut sub numele de *plasticitate a creierului*, tăria unor asemenea legături se poate schimba de la un moment la altul, poate chiar pentru timpi mai mici de o secundă, aşa cum se pot schimba chiar conexiunile însele. Se pune atunci întrebarea importantă: care sunt procedurile care guvernează aceste schimbări sinaptice?

În modelele conexioniste, (aşa cum au fost ele adaptate pentru reţele neuronale artificiale) există un anumit tip de *regulă computaţională* care guvernează schimbările sinaptice. Această regulă este specificată astfel încât sistemul poate să-şi îmbunătăţească performanţa trecută pe baza unor criterii formulate în prealabil în relaţie cu *inputurile* sale externe. Deja în 1949, Donald Hebb a sugerat o regulă simplă de acest fel. Modelele conexioniste moderne [1] au modificat considerabil procedura Hebb originară, în diferite feluri. În modelele generale de acest tip, trebuie să existe o regulă computaţională bine definită deoarece modelele sunt ele însele lucruri care pot fi rulate pe un calculator obişnuit (conform paragrafului 1.5). Dar esenţa a ceea ce am spus în partea întâi este că nici o astfel de procedură computaţională nu poate fi adecvată pentru a explica toate manifestările operaţionale ale înţelegerii umane conştiente. Astfel, trebuie să căutăm ceva diferit pentru a da seama de tipul potrivit de 'mecanism de control' cel puţin în cazul schimbărilor sinaptice care pot avea o oarecare relevanţă asupra activităţii *conştiente* actuale.

Au fost sugerate şi alte idei, ca de exemplu cele ale lui Gerald Edelman în recenta sa carte *Bright Air, Brilliant Fire* (1992) (şi precedenta sa trilogie; Edelman 1987,1988,1989), în care propune în locul unor reguli de tip hebbian, o formă de principiu 'darwinist' care să opereze asupra creierului, permiţându-i să-şi îmbunătăţească în mod continuu performanţele prin intermediul unui tip de selecţie naturală care guvernează aceste conexiuni. Există în acest model asociaţii semnificative cu felul în care sistemul imunitar îşi dezvoltă posibilităţile de 'recunoaştere' a substanţelor. Centrul de greutate al descrierii se mută acum asupra rolului complicat al neurotransmiţătorului şi a altor chimicale care sunt

implicate în comunicarea dintre neuroni. Totuşi, aceste procese, aşa cum sunt ele concepute în prezent, sunt încă tratate într-un mod clasic şi computaţional. Într-adevăr, Edelman şi colegii săi au construit o serie de aparate controlate computaţional (numite DARWIN I, II, III, IV, etc.) care intenţionează să simuleze, într-un mod din ce în ce mai complex, chiar tipurile de proceduri pe care le atribuie acţiunii mentale. Din faptul că un calculator obişnuit este cel care efectuează operaţiunea de control, rezultă că această schemă particulară este, încă, una computaţională -- cu un anumit sistem particular de reguli *bottom-up*. Nu contează cât de diferită este, în detaliu, o astfel de schemă de alte proceduri computaţionale; ea este cuprinsă în aceeaşi categorie cu cele din partea întâi, (conform paragrafelor 1.5, 3.9 şi argumentele din dialogul fantastic din paragraful 3.23). Datorită acestor argumente este extrem de improbabil ca orice lucru de acest tip să poată furniza un model actual al minţii conştiente.

Pentru a scăpa de aceste constrângeri computaţionale avem nevoie de alte tipuri de control a conexiunilor sinaptice. Un astfel de control ar trebui să cuprindă un proces fizic în care o anumită formă de coerenţă cuantică să aibă un rol important de jucat. Dacă acest proces este similar într-un mod esenţial cu acţiunea sistemului imunitar, atunci sistemul imunitar însuşi trebuie să depindă de efectele cuantice. Poate că există într-adevăr ceva esenţialmente cuantic în felul în care mecanismele de recunoaştere ale sistemului imunitar operează -- aşa cum a susţinut, în particular, Michael Conrad, (1990,1992, 1993). Acest lucru nu m-ar surprinde, dar rolul acţiunii cuantice în funcţionarea sistemului imunitar nu formează o parte importantă a modelului Eldeman al creierului.

Chiar dacă conexiunile sinaptice sunt controlate în vreun fel de efecte cuantice coerente, este dificil de văzut dacă poate exista ceva esenţial cuantic legat de activitatea de semnalizare nervoasă propriu-zisă. Cu alte cuvinte, este greu de văzut cum anume poate cineva să conceapă o superpoziţie cuantică efectivă între un neuron care emite şi nu emite, în acelaşi timp. Semnalele nervoase par destul de macrosopice pentru a ne convinge că este greu de crezut într-o astfel de imagine, în ciuda faptului că transmisiunea nervoasă este mai curând bine izolată de prezenţa tecii de mielină care înconjoară nervul. Conform punctului de vedere OR pe care l-am susţinut în paragraful 6.12, ar trebui să ne aşteptăm ca reducţia de stare obiectivă să aibă loc rapid, atunci când un neuron emite, nu pentru că ar exista o mişcare de masă la scară mare (nu la standardele cerute, cel puţin), ci

datorită existenţei unui câmp electric care se propagă de-a lungul nervului, câmp cauzat de semnalul nervos, şi suficient de puternic pentru a putea fi în principiu detectabil în zona imediat înconjurătoare. Acest câmp va perturba, în mod aleator, cantităţi destul de mari din materialul cerebral înconjurător — destul, poate, pentru satisfacerea criteriului din paragraful 6.12 aproape imediat ce semnalul este transmis. Astfel, menţinerea superpoziţiilor cuantice ale neuronului care emite şi nu emite în acelaşi timp pare implauzibilă.

7.3. Calculabilitate cuantică

Această proprietate a neuronului emitent de a perturba mediul este trăsătura cea mai inadecvată tipului de propunere explicativă generală în favoarea căreia am argumentat în **MN**, în care aveam nevoie de un tip de superpoziţie cuantică a neuronului care să emită şi să nu emită în acelaşi timp. Cu actualul criteriu **OR** pentru reducţia de stare, procesului **R** îi corespunde o perturbare mai mică a mediului decât era cazul mai înainte, şi este greu de crezut în posibilitatea ca menţinerea unor asemenea superpoziţii să fie semnificativă. Ideea era că dacă ar fi fost posibilă efectuarea mai multor 'calcule' separate asupra diferitelor forme de superpoziţie, atunci creierul ar dobândi ceva din natura unui *calcul cuantic*, în locul simplului calcul Turing. În ciuda aparentei implauzibilităţi a operativităţii calculului cuantic la acest nivel al activităţii cerebrale, va fi util să examinăm anumite aspecte privind această noţiune.

Calculul cuantic este un concept teoretic propus iniţial de David Deutsch (1985) şi Richard Feynman (1985,1986) vezi de asemenea Benioff (1982), Albert (1983)- şi este în prezent explorată de un grup mai larg. Ideea este de a extinde noţiunea clasică de maşină Turing în domeniul cuantic, la o maşină Turing cuantică. Toate operaţiile pe care această 'maşină' extinsă le efectuează sunt subordonate legilor cuantice — cu superpoziţii permise- care se aplică la sistemele de nivel cuantic.

Astfel, pentru cea mai mare parte a evoluţiei dispozitivului, acţiunea lui **U** este predominantă, iar prezervarea superpoziţiilor este o parte esenţială a acţiunii sale. Procedura **R** ar deveni relevantă doar la *sfârşitul* operaţiei, atunci când sistemul este 'măsurat', pentru a formula rezultatul calculului. De fapt, chiar dacă

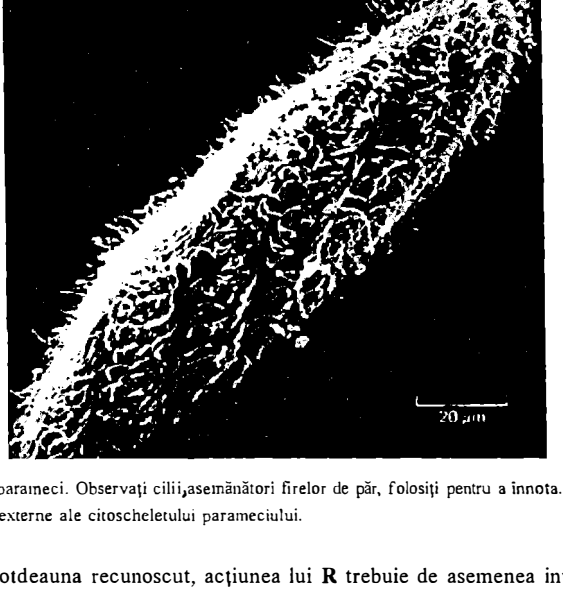


Fig. 7.2. Un parameci. Observaţi cilii,asemănători firelor de păr, folosiţi pentru a înnota. Ei formează extremităţile externe ale citoscheletului parameciului.

nu este întotdeauna recunoscut, acţiunea lui **R** trebuie de asemenea invocată din timp în timp într-un fel secundar, în decursul calculului, pentru a stabili dacă s-a terminat sau nu.

S-a descoperit că, deşi un calculator cuantic nu poate obţine nimic din ce n-a fost deja obţinut, *în principiu* de către o maşină Turing convenţională, există anumite clase de probleme pentru care calculul cuantic este capabil să depăţească calculul Turing în sensul *teoriei complexităţii* (vezi Deutsch, 1985). Pentru aceste clase de probleme calculatorul cuantic este, în principiu mult mai rapid decât cel convenţional (vezi, de exemplu, Deutsch şi Josza (1992) pentru o clasă de probleme interesante, deşi oarecum artificiale, în care calculatorul cuantic excelează). Conform argumentaţiei recente a lui Peter Shor, calculul cuantic poate rezolva (în timp polinomial) problema interesantă a factorizării întregilor mari.

În calculul cuantic 'standard', sunt adoptate regulile uzuale ale teoriei cuantice, în care sistemul evoluează conform procedurii **U** pentru întreaga operaţie, cu **R** apărând doar în anumite locuri. Nu e nimic 'necalculabil' în această

procedură, în sensul obişnuit al 'calculabilului', deoarece U este o operaţie calculabilă, iar **R** este un procedeu pur probabilistic. Ceea ce se poate obţine, în principiu, cu ajutorul unui calculator cuantic, poate fi de asemenea obţinut, în principiu, de o maşină Turing potrivită, înzestrată cu un algoritm de producere de numere aleatoare. Astfel, conform argumentelor din partea întâia, nici măcar un calculator cuantic nu este în stare să efectueze operaţiile înţelegerii umane conştiente. O speranţă ar fi ca subtilităţile privind ceea ce se petrece *cu adevărat* atunci când vectorul de stare pare să fie 'redus', să ne ducă la ceva *cu adevărat* necalculabil, în locul procedurii **R** de producere de opriri bruşte aleatoare Astfel, teoria completă a procesului presupus **OR** va trebui să fie construită pe o schemă *esenţial necalculabilă.*

Ideea din **MN** era de a privi acţiunea suprapusă a calculelor Turing ca fiind periodic întreruptă de o acţiune necalculabilă care nu poate fi înţeleasă decât dacă avem la îndemână o nouă fizică care să intre în acţiune ori de câte ori apare **R**. Dacă însă astfel de superpoziţii de calcule ne sunt interzise în cazul neuronilor pentru că o prea mare parte din mediul înconjurător al neuronului este perturbat de semnalul nervos, este greu de văzut cum ar putea fi posibilă întrebuinţarea ideii de calcul cuantic standard, ca să nu mai vorbim de orice modificare a procedeului în favoarea unei înlocuiri al lui **R** cu ceva non-computaţional, ca **OR**. Cu toate acestea, vom vedea în curând că avem la dispoziţie o altă posibilitate, mult mai promiţătoare. Dar pentru a înţelege despre ce e vorba, va trebui să aruncăm o privire asupra naturii biologice a celulelor creierului.

7.4. Citoschelete şi microtubule

Dacă ar fi să credem că neuronii sunt singurele lucruri care controlează acţiunile sofisticate ale animalelor, umilul parameci ne-ar pune o problemă profundă: un animal care înnoată în mica lui băltoacă cu numeroasele sale picioruşe păroase, numite cili, în direcţia hranei (alcătuită din bacterii), retrăgându-se la apariţia unei primejdii, gata să înnoate într-o altă direcţie. Parameciul poate, de asemenea, să perceapă obstacolele şi să le ocolească. Mai mult, pare să poată chiar *învăţa* din experienţele trecute [2] deşi această facultate, cea mai remarcabilă, îi este întrucâtva contestată [3]. Cum pot fi asemenea performanţe realizate de către

un animal fără nici un singur neuron? Într-adevăr, fiind un organism monocelular (şi nefiind el însuşi un neuron), parameciul nu are cum să dezvolte asemenea accesorii complicate (vezi figura 7.2)

Şi totuşi, trebuie să existe un sistem complicat de control care guvernează comportamentul parameciului sau al altor animale monocelulare, ca amoeba. Însă acest sistem de control nu este un sistem nervos. Structura responsabilă cu aceste funcţii este, aparent, o parte a ceea ce se numeşte *citoschelet*. Aşa cum numele o sugerează, citoscheletul furnizează cadrul care păstrează forma celulei; el face însă mult mai mult decât atât. Cilii înşişi sunt terminaţiii ale fibrelor citoscheletice; în plus, citoscheletul pare să conţină sistemul de control al celulei; şi asta pe lângă atribuţiile privind 'ţinerea împreună' şi transportul diferitelor molecule dintr-o parte intr-alta. Pe scurt, citoscheletul pare să joace rolul combinat de schelet, sistem muscular, picioare, sistem circulator şi nervos, pentru un organism monocelular!

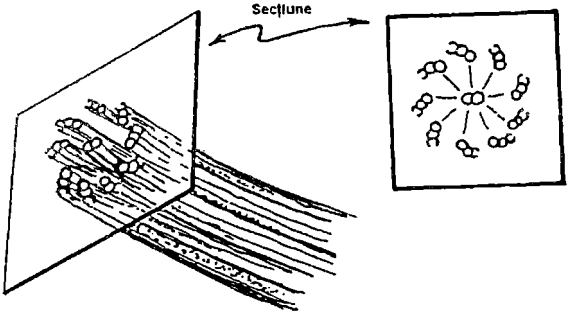


Fig. 7.3. Părţi importante ale citoscheletului consistă în legături de mici tuburi, organizate într-o structură cu secţiunea transversală ca un ventilator. Cilii parameciului sunt astfel de legături.

În cele ce urmează, ne va interesa rolul de 'sistem nervos' al citoscheletului. Cum neuronii noştri sunt celule individuale, putem spune că fiecare din ei are *propriul* său citoschelet. Înseamnă oare aceasta că există un sens în care fiecare neuron individual poate avea ceva de genul unui 'sistem nervos personal'? E o problemă curioasă, şi există un număr de cercetători care au ajuns la concluzia că ceva de acest gen ar putea fi adevărat (vezi cartea de deschizător de drumuri a lui Stuart Hameroff din 1987, *Ultimate computing:Biomolecular Consciousness and*

Nano-Technology; de asemenea, Hameroff şi Watt (1982), precum şi numeroase articole din revista *Nanobiology*).

Pentru a discuta asemenea probleme, trebuie să aruncăm mai întâi o privire asupra felului în care este alcătuit citoscheletul. Acesta constă din molecule de tipul proteinelor, aranjate în diferite tipuri de structuri: actini, microtubule şi filamente intermediare. Pe noi ne vor interesa în special microtubulele: tuburi cilindrice de aproximativ 25nm în diametru pe dinafară şi 14nm pe dinăuntru (unde un nanometru, 'nm' este 10^{-9} m), organizate uneori în fibre de forma tubulară mai mari, alcătuite din nouă dubleţi, tripleţi sau tripleţi parţiali de microtubule, dispuse într-un aranjament a cărui secţiune transversală aminteşte un ventilator (vezi figura 7.3), uneori cu o pereche de microtubule în centru. Cilii parameciului sunt structuri de acest tip.

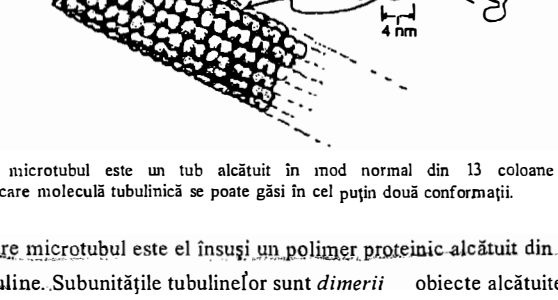


Fig. 7.4. Un microtubul este un tub alcătuit în mod normal din 13 coloane de dimeri tubulinici. Fiecare moleculă tubulinică se poate găsi în cel puţin două conformaţii.

Fiecare microtubul este el însuşi un polimer proteinic alcătuit din subunităţi numite tubuline. Subunităţile tubulinelor sunt *dimerii* obiecte alcătuite din două părţi separate numite α-tubuline şi β-tubuline, iar fiecare din acestea este compusă din aproximativ 450 de aminoacizi. Tubulinele sunt perechi de proteine globulare, de forma unei arahide, organizate într-o reţea oarecum hexagonală de-a lungul întregului tub, aşa cum se poate vedea în figura 7.4. În general sunt 13 coloane de dimeri tubulinici pentru fiecare microtubul. Fiecare dimer este de aproximativ 8nm $\times$ 4nm $\times$ 4nm, iar numărul său atomic este de aproximativ 11×10^4 (ceea ce înseamnă că are cam acest număr de nucleoni, astfel încât masa sa, în unităţi absolute de masă ar fi cam de 10^{-14}).

Fiecare dimer tubulinic, ca întreg, poate exista în cel puţin două configuraţii

geometrice diferite, numite *conformaţii* diferite. În una din acestea, dimerii sunt înclinaţi într-o direcţie de aproximativ 30% în raport cu direcţia microtubulului. Există dovezi care arată că aceste două conformaţii diferite corespund celor două stări diferite ale polarizării electrice a dimerului care se datorează, la rândul lor, faptului că un electron, plasat central la joncţiunea dintre tubulina-α şi tubulina-β, poate trece dintr-o poziţie într-alta.

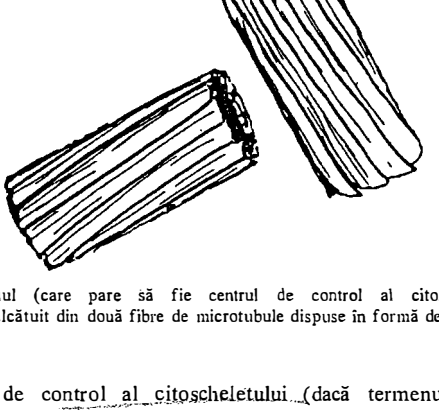


Fig. 7.5. Centriolul (care pare să fie centrul de control al citoscheletului, dacă aşa ceva există) este alcătuit din două fibre de microtubule dispuse în formă de "T" (separat), de tipul celor din Fig.7.3.

Centrul de control al citoscheletului (dacă termenul este într-adevăr corespunzător), este o structură numită *centriol*. Acesta pare să fie alcătuit din doi cilindri de nouă tripleţi de microtubule, cei doi cilindri alcătuind o structură în formă de 'T' (figura 7.5) (cilindrii sunt similari, în linii mari, celor care apar în cilii parameciului, sau în figura 7.3). Centriolul formează partea critică a unei structuri numită *centrul de organizare al microtubulelor* sau *centrozom*. Oricare ar fi rolul centriolului în timpul existenţei normale a unei celule, el are cel puţin o sarcină fundamentală; într-un moment critic, fiecare din cei doi cilindri ai centriolului produc un alt cilindru, alcătuind doi centrioli în formă de 'T', separaţi unul de altul. Fiecare dintre aceştia pare să ducă cu sine o legătură de microtubule sau, mai riguros vorbind, fiecare devine un centru în jurul căruia microtubulele se vor asambla. Aceste fibre de microtubule conectează cumva centriolul cu lanţurile de ADN din nucleu (în punctele lor centrale, numite centromeri) —iar lanţurile de ADN se despart, iniţiind acel proces extraordinar numit tehnic *mitoză*, care înseamnă pur şi simplu *diviziune celulară* (vezi figura 7.6)

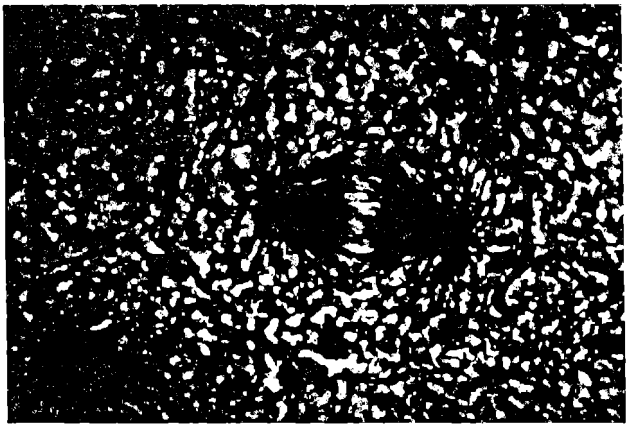


Fig. 7.6. În mitoză (diviziune celulară) cromozomii sunt separați cu ajutorul microtubulelor.

Poate părea ciudat să existe două centre de comandă diferite într-o singură celulă. Pe de-o parte, nucleul unde se găsește materialul genetic fundamental al celulei, care controlează ereditatea și identitatea ei particulară și guvernează producerea de material proteinic din care este compusă celula; pe de altă parte, *centrozomul*, cu componenta lui cheie, *centriolul*, punctul focal al citoscheletului, pare să controleze mișcările celulei și organizarea ei de detaliu. Se crede că prezența acestor două structuri diferite în celulele eucariote (adică celulele aproape tuturor animalelor și plantelor de pe această planetă, cu excepția bacteriilor, algelor verzi-albastre și a virușilor) este rezultatul unei 'infecții' care a avut loc cu mii de milioane de ani în urmă. Celulele care locuiau pământul atunci erau celulele procariote care mai există astăzi sub formă de bacterii sau alge verzi-albastre și care nu au citoschelet. Una dintre sugestii (Sagan, 1976) este că anumite procariote timpurii s-au unit sau au fost 'infectate' cu un anumit tip de spirochete, un organism care înnoată cu ajutorul unei cozi compuse din proteine citoscheletice. Aceste două tipuri diferite de organisme au învățat să trăiască împreună într-o relație simbiotică, ca o singură celulă eucariotă. Astfel, spirochetele au devenit în cele din urmă citoscheletele celulei cu toate implicațiile corespunzătoare pentru evoluția viitoare care, în cele din urmă ne-a făcut posibilă apariția!

Organizarea microtubulelor mamiferelor este interesantă dintr-un punct de

vedere matematic. Numărul 13 nu pare să aibă, la prima vedere, vreo semnificaţie matematică particulară, însă acest lucru nu este chiar aşa. El este unul dintre faimoasele numere ale lui Fibonacci:

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144,...

în care fiecare număr succesiv este obţinut ca o sumă a celor două numere precedente. Această observaţie ar putea părea neinteresantă; se ştie însă, numerele Fibonacci sunt nişte numere care apar foarte frecvent (de obicei la scară mult mai mare) în sistemele biologice.

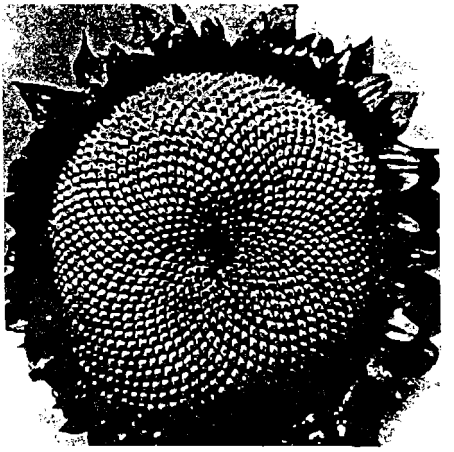


Fig. 7.7. Floarea soarelui. La multe plante, intervin numerele Fibonacci. În regiunile exterioare sunt 89 spirale în sensul acelor de ceasomic şi 55 în sens invers. În zona centrală găsim alte numere Fibonacci.

De exemplu, în conurile de pin, mijlocul florii soarelui, sau în trunchiurile de palmieri, se găsesc aranjamente spirale sau elicoidale cuprinzând interpenetrarea unor curbe orientate spre dreapta sau spre stânga, cu numărul de 'rânduri', orientate spre dreapta sau spre stânga, reprezentând două numere Fibonaci succesive (vezi figura 7.7). Dacă examinăm structurile pornind simultan de la cele două capete, putem găsi un loc de 'creştere'; cele două numere se deplasează atunci spre următoarea pereche Fibonacci. Destul de curios, forma hexagonală a microtubulelor prezintă o trăsătură foarte asemănătoare uneori, cu o

organizare chiar mai precisă şi s-a găsit, se pare, că această formă este făcută din 5 aranjamente elicoidale orientate dextrogir şi 8 orientate levogir, aşa cum sunt reprezentate în figura 7.8. În figura 7.9 am încercat să arăt cum am putea să considerăm această structură ca ceva 'văzut' din interiorul unor microtubule. Numărul 13 apare aici ca o sumă 5+8. Este de asemenea curios că microtubulele duble care apar frecvent par să aibă în mod normal 21 de coloane de dimeri tubulinici care formează frontiera exterioară a tubului compus — următorul număr Fibonacci! (Totuşi, nu trebuie să vă lăsaţi duşi de aceste consideraţii; de exemplu numărul 9, care apare în fibrele de microtubule din cili şi centrioli *nu* este un număr Fibonacci).

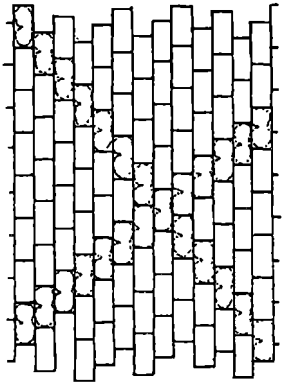


Fig. 7.8. Imaginaţi-vă un microtubul tăiat în lungime şi desfăşurat de-a lungul tăieturii. Tubulinele apar ordonate în linii înclinate ce se întâlnesc la al cincilea sau al optulea element.

De ce apar numerele Fibonacci în structura microtubulelor? În cazul conurilor de pin, al floarii soarelui, etc., există diferite teorii plauzibie Alan Turing a fost şi el printre cei ce s-au gândit serios la acest subiect (Hodges, 1983,p.437). S-ar putea ca aceste teorii să nu fie deloc potrivite pentru microtubule, unde ne-am aştepta ca alt gen de idei să fie relevante. Koruga (1974) a sugerat că aceste numere Fibonacci ar putea furniza avantaje pentru microtubul în calitatea sa de 'procesor de informaţie' Într-adevăr, Hameroff şi colegii săi [4], în urmă cu mai mult de zece ani, au argumentat că microtubulele ar putea juca rolul de *automate celulare*, în care se pot transmite şi procesa de-a lungul tuburilor, semnale complicate, sub

formă de unde de stări de polarizare ale tubulinelor. Starea fiecărui dimer va fi influenţată de stările de polarizare a şase dintre vecinii săi (din cauza interacţiunilor Van der Waals dintre ei) dând naştere la anumite reguli specifice care guvernează conformaţia fiecărui dimer în funcţie de conformaţia vecinilor săi. Acest lucru ar permite propagarea tuturor tipurilor de mesaje de-a lugul fiecărei microtubule. Aceste semnale par să fie relevante pentru felul în care microtubulele transportă diferite molecule de-a lungul lor, şi pentru diferitele conexiuni între microtubule învecinate de forma unor punţi de proteine numite MAP (proteine asociate microtubulelor – vezi figura 7.10). Koruga susţine că există un grad sporit de eficienţă în cazul structurii care cuprinde numere Fibonacci, ca în cazul observat pentru microtubule. Trebuie, într-adevăr să existe motive serioase pentru acest tip de organizare a microtubulelor, deoarece, deşi există o anume variaţie a numerelor care se aplică celulelor eucariote în general, alcătuirea cu 13 coloane pare să fie aproape universală printre mamifere.

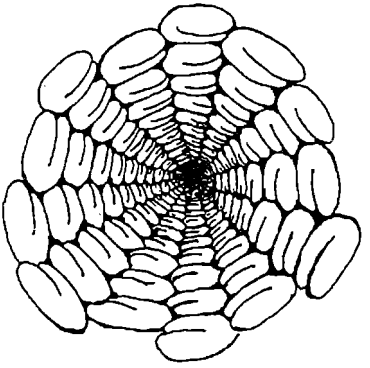


Fig. 7.9. Vedere înspre interiorul unui microtubul. Se poate observa aranjamentul spiral 5+8 al tubulinelor.

Care este semnificaţia microtubulelor pentru neuroni? Fiecare neuron individual are propriul său citoschelet. Dar care este rolul acestuia? Sunt sigur că mai sunt multe de cercetat în acest domeniu, însă avem chiar şi până acum un număr considerabil de fapte cunoscute. În particular, microtubulele din neuroni pot fi foarte lungi în comparaţie cu diametrul lor (care este numai de aproximativ 25

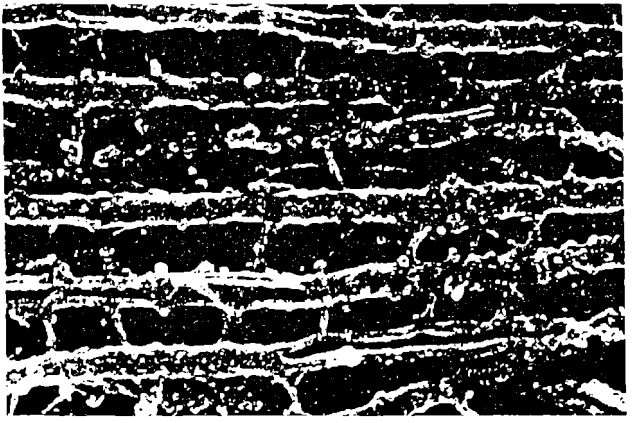


Fig. 7.10. Microtubulele tind să fie interconexate cu cele învecinate prin punți de proteine asociate microtubulelor (MAP)

30nm) pot atinge câțiva milimetri, sau chiar mai mult. În plus, ele se pot lungi sau contracta, după împrejurări, și pot transporta molecule neurotransmițătoare. Există microtubule de-a lungul axonilor și al dendritelor. Cu toate că microtubulele individuale nu par să se extindă pe întreaga lungime a unui axon, ele formează cu siguranță rețele care au această proprietate, fiecare microtubul comunicând cu următorul prin intermediul unei punți de proteine (MAP). Microtubulele par să fie răspunzătoare pentru menținerea tăriei sinapselor și, fără îndoială, pentru alterarea acestora atunci când este nevoie. Mai mult, ele par să organizeze creșterea terminațiilor nervoase, ghidându-le spre conexiunile lor cu alte celule nervoase.

Cum neuronii nu se mai divid după ce creierul este complet format, nu există nici un rol în neuron pentru centrioli. Într-adevăr, centriolii par să fie absenți din centrozomul neuronului care se află aproape de nucleul neuronului. Microtubulele se extind de aici până în vecinătatea terminațiilor presinaptice ale axonului și, de asemenea, în cealaltă direcție, în dendrite și în prelungirile dendritice care formează de obicei capătul post-sinaptic al unei punți sinaptice (figura 7.12). Aceste prelungiri sunt supuse creșterii și degenerării, un proces care pare să formeze o parte importantă a plasticității creierului, în care o importantă parte a interconexiunilor generale din creier suferă schimbări subtile continue. Par să existe

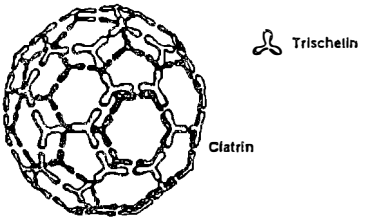


Fig. 7.11. O moleculă clatrin (similară ca structură globală unei fulerene, dar constituită din substructuri mai complicate – proteine în loc de atomi de carbon).

dovezi semnificative privind faptul că microtubulele sunt într-adevăr implicate într-un mod esenţial în controlul plasticităţii creierului.

Ca o curiozitate poate fi de asemenea menţionat faptul că în terminaţile presinaptice ale axonilor există anumite substanţe asociate cu microtubulele, fascinante din punct de vedere geometric şi importante pentru eliberarea neurotransmiţătorilor chimici. Aceste substanţe, numite *clatrini*, sunt construite din trimeri proteinici cunoscuţi sub denumirea de *triskelioni clatrinici* care formează structuri arborescente (polipeptidice). Triskelioni clatrinici împreună formează frumoase configuraţii matematice identice în organizarea generală cu moleculele de carbon numite *fulerene* (sau 'mingi *bucky*') datorită similitudinilor cu faimoasele domuri geodezice construite de arhitectul american, Buckminster Fuller [5]. Clatrinii sunt totuşi mult mai mari decât moleculele fulerene; un întreg triskelion clatrinic, o structură care conţine câţiva aminoacizi, ia locul unui singur atom de carbon dintr-o astfel de 'minge'. Clatrinii particulari care răspund de eliberarea substanţelor neurotransmiţătoare la sinapse par să aibă în pricipal o structură de icosaedru (vezi figurile 7.11 şi 7.12).

În secţiunea precedentă, întrebarea importantă a fost: ce anume guvernează variaţia de tărie a sinapselor şi organizează locurile în care vor fi făcute conexiunile sinaptice în funcţiune?

Iată că am fost conduşi la ideea clară că citoscheletul este cel care trebuie să joace un rol central în acest proces. Cum ne poate ajuta asta în căutarea unui aspect non-computaţional al minţii? Până acum, tot ce am câştigat pare să fie doar o mare creştere potenţială a puterii de calcul, mult peste şi dincolo de ceea ce am fi avut dacă unităţile de lucru ar fi fost doar neuronii individuali.

Într-adevăr, dacă dimerii tubulinici sunt unităţile computaţionale de bază, atunci trebuie să luăm în considerare posibilitatea unei puteri de calcul a creierului care se extinde mult dincolo de cea presupusă până acum în inteligenţa artificială. Hans Moravec, în cartea sa *Mind Children* (1988) presupune, pe baza unui model care ia în considerare doar neuronii, că mintea umană ar putea, în principiu, atinge viteza de 10^{14} operaţii de bază pe secundă, dar nu mai mult – considerând în acelaşi timp că avem 10^{11} neuroni operaţionali, fiecare capabil să trimită 10^3 semnale pe secundă (vezi paragraful 1.2).

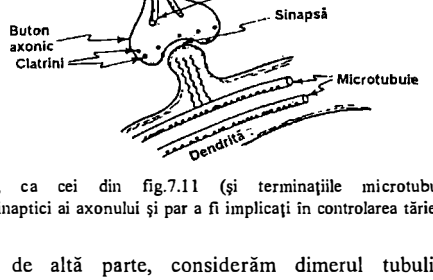


Fig. 7.12. Clatrinii, ca cei din fig.7.11 (şi terminaţiile microtubulilor), se află în interiorul butonilor sinaptici ai axonului şi par a fi implicaţi în controlarea tăriei sinapsei.

Dacă, pe de altă parte, considerăm dimerul tubulinic ca unitatea computaţională de bază, trebuie să luăm în consideraţie că există aproximativ 10^4 dimeri pe neuron, operaţiile elementare pot fi efectuate acum de 10^6 ori mai repede, ceea ce ne dă un total de aproximativ 10^{24} operaţii pe secundă[5] Calculatoarele actuale pot să se apropie abia de primul dintre aceste numere, cel de 10^{14} operaţii pe secundă, aşa cum susţin cu tărie Moravec şi alţii, neîntrevăzându-se încă nici o speranţă ca viteza de 10^{24} operaţii pe secundă să poată fi luată în considerare într-un viitor previzibil.

Desigur, se poate susţine cu bune argumente că nu poate fi vorba în nici un caz de o operare a creierului cu o eficienţă microtubulară de 100% aşa cum presupun aceste cifre. Cu toate acestea, este clar că posibilitatea unui 'calcul

---

[5] Faţă de aceste valori, din ediţia iniţială (Oxford University Press Inc., New York), care stă la baza acestei traduceri, în cartea de la Vintage Press, apărută câteva luni mai târziu, apar valori cu trei ordine de mărime mai mari, deci 10^{27} respectiv 10^{17} în loc de 10^{24} şi 10^{14}, fără nici o referire la această diferenţă sensibilă, chiar dacă mult peste posibilităţile de calcul (n.t.)

microtubular' (conform cu Hameroff 1987) pune într-o perspectivă complet diferită câteva dintre argumentele privind iminenţa construirii unei inteligenţe artificiale de calibrul celei umane. Se pune chiar întrebarea dacă mai putem accepta sugestiile [6] conform cărora facultăţile mentale ale unui vierme nematod au fost atinse pentru că organizarea sa neurală a fost cartografiată şi simulată computaţional.

Aşa cum am remarcat în paragraful 1.15, capacităţile actuale ale unei furnici par să depăşească cu mult tot ce a fost realizat până acum folosind procedurile standard ale inteligenţei artificiale.

Ne putem întreba cât de mult câştigă o furnică din enorma cantitate de 'procesori de informaţie microtubulari', în opoziţie cu cazul în care ar avea doar 'întrerupători neuronali'. Ca şi în cazul paramecilor, nici pentru furnici nu avem încă un răspuns.

Şi totuşi, argumentele din partea întâia a acestei cărţi, pretind ceva mai mult. Eu susţin că facultăţile intelectului uman se află dincolo de orice schemă computaţională posibilă. Dacă microtubulele sunt cele care controlează activitatea creierului, atunci trebuie să existe ceva la nivelul de acţiune al microtubulelor fundamental diferit de simplul calcul. Am argumentat de ce o astfel de acţiune necalculabilă trebuie să fie rezulatul unui fenomen cuantic coerent la scară rezonabil de mare, cuplat într-un anumit fel cu comportamentul macroscopic, astfel încât sistemul să fie în stare să profite de pe urma acelui nou proces care va lua locul procedurii **R** a fizicii zilelor noastre. Iar ca un prim pas, trebuie să căutăm un loc pentru *coerenţa cuantică* în activitatea citoscheletică.

7.5. Coerenţă cuantică în microtubule?

Există vreo dovadă pentru a presupune acest lucru? Să ne reamintim ideile lui Fröhlich (1975) privind posibilitatea unor fenomene de coerenţă cuantică în sistemele biologice la care ne-am referit în paragraful 7.1. Fröhlich argumenta că, atâta timp cât energia metabolică este destul de mare, iar proprietăţile dielectrice ale materialelor considerate sunt suficient de diferite, există posibilitatea unei coerenţe cuantice la scară mare, similară cu cea care apare în fenomenele de superconductivitate şi superfluiditate numite uneori condensarea Bose-Einstein –

chiar la temperaturile destul de mari care există în sistemele biologice. Iată că suntem în măsură să observăm nu numai că energia metabolică este destul de mare, iar proprietăţile dielectrice neobişnuit de diferite (un fapt observaţional tulburător care l-a determinat în anii treizeci pe Fröhlich să propună această linie de gândire), dar există acum şi dovezi experimentale directe privind oscilaţiile cu frecvenţa de 10^{11} Hz din interiorul celulelor, observaţii cuprinse în predicţiile lui Fröhlich (vezi Grundler şi Keilmann, 1983).

Într-un condensat Bose-Einstein (care apare de asemenea într-un sistem laser), un număr mare de particule participă colectiv la o singură stare cuantică. Există o funcţie de undă pentru această stare, de tipul celei pentru o singură particulă, dar care se aplică acum întregii colecţii de particule care participă la starea cuantică. Să ne reamintim natura contra-intuitivă a stării (difuzate) cuantice a unei singure particule (paragrafele 5.6, 5.11). Într-un condensat Bose-Einstein lucrurile se întâmplă ca şi cum întregul sistem, conţinând un mare număr de particule, se comportă ca un întreg, aproape la fel cum s-ar comporta o particulă individuală, cu excepţia unei scalări corespunzătoare. Există o coerenţă la scară mare; multe dintre trăsăturile stranii ale funcţiei de undă cuantice se păstrază la nivel macroscopic.

Ideile originale ale lui Fröhlich păreau să susţină faptul că astfel de stări cuantice ar putea apărea în membranele celulelor[6]; acum, însă, ni se prezintă posibilitatea adiţională şi poate mai plauzibilă conform căreia în microtubule trebuie să căutăm o comportare cuantică de acest tip. Par să existe mai multe dovezi în favoarea acestui punct de vedere [7]. Încă din 1974, Hameroff (1974) a propus ca microtubulele să acţioneze ca 'ghiduri de undă dielectrice'. Este într-adevăr tentant să credem că Natura a ales aceste tuburi în structurile citoscheletice cu un bun motiv. Poate tuburile însele servesc la furnizarea izolării efective care permite stării cuantice din interiorul tubului să rămână separată de mediul înconjurător pentru o perioadă apreciabilă de timp. Este interesant de observat, în acest sens, că Emilio del Giudice şi colegii săi de la Universitatea din Milano (del Giudice *et.al.*,1983)

---

[6] Un puternic susţinător al ideii după care condensarea Bose–Einstein ar putea oferi „sensul unităţii sinelui" care pare caracteristica conştiinţei, în relaţie cu ideile lui Frohlich este Ian Marshall (1989). Vezi de asemenea, Zohar (1990), Zohar şi Marshall (1994) şi Lockwood (1989). Un susţinător mai vechi al ideii de activitate cuantică globală coerentă de tip 'holografic' în creier a fost Karl Pribram (1966, 1975, 1991).

au argumentat că un efect cuantic auto-focalizat al undelor electromagnetice în interiorul materialului citoplasmatic din celule provoacă semnale care pot fi confinate la dimensiuni de ordinul diametrului intern al microtubulelor. Acest lucru poate susţine substanţial teoria ghidului de undă, dar efectul poate fi de asemenea efectiv în chiar formarea microtubulelor.

Mai există un lucru de interes aici, şi acesta priveşte chiar natura *apei*. Tuburile par să fie goale – un fapt curios şi poate semnificativ el însuşi, dacă considerăm că aceste tuburi ne furnizează condiţiile controlate favorabile unui anumit tip de oscilaţii cuantice colective. 'Gol' înseamnă, în acest caz, că ele conţin doar apă (fără alţi ioni dizolvaţi în ea). Ne putem deci închipui că apa, cu moleculele sale mişcându-se aleator, nu este un tip de structură suficient de organizată pentru ca apariţia oscilaţiior cuantice coerente să fie probabilă. Cu toate acestea, apa care este găsită în celule nu seamănă de loc cu apa obişnuită – dezordonată, cu moleculele mişcându-se într-un mod incoerent şi complet la întâmplare. O parte din ea (şi cât anume, asta-i o problemă controversată) există într-o stare *ordonată* (numită uneori apă vicinală, vezi Hameroff, 1987, p.172). O astfel de stare ordonată a apei se poate extinde cale de 3nm sau chiar mai mult în afara suprafeţelor citoscheletului. Nu pare nerezonabil să presupunem atunci că apa din interiorul microtubulelor este de asemenea de natură ordonată, iar acest lucru ar favoriza puternic posibilitatea unor oscilaţii cuantice în interiorul, sau în relaţie cu aceste tuburi (vezi, în particular, Jibu et.al, 1994).

Oricare ar fi statutul ultim al acestor idei interesante, un lucru pare clar: este puţin probabil ca o discuţie clasică a citoscheletului să-i poată explica comportamentul. Acelaşi lucru s-ar putea spune chiar despre neuroni, unde discuţia în termeni clasici pare de asemenea nepotrivită. O simplă examinare a literaturii care se ocupă de acţiunea citoscheletului ne arată recursul permanent la conceptele mecanicii cuantice şi nu am nici o îndoială că acest gen de abordare va fi din ce în ce mai popular.

Cu toate acestea, este clar de asemenea că există mulţi cercetători încă neconvinşi de posibilitatea unor efecte cuantice semnificative relevante la nivelul citoscheletului sau acţiunii cerebrale. Chiar dacă ar exista efecte importante de natură cuantică esenţiale pentru funcţionarea microtubulelor şi pentru funcţionarea creierului, nu este uşor să le pui în evidenţă prezenţa printr-un experiment definitiv.

Dacă suntem norocoşi, unele dintre procedurile standard care servesc la demonstrarea prezenţei condensatului Bose-Einstein în sistemele fizice – ca, de exemplu, în supraconductivitatea la temperaturi mari – ar putea fi de asemenea aplicabile în cazul microtubulelor. Pe de altă parte, s-ar putea să nu fim chiar atât de norocoşi şi să trebuiască să ne gândim la ceva nou.

O posibilitate interesantă ar fi să demonstrăm că excitaţiile microtubulelor prezintă tipul de nonlocalitate care apare în fenomenele EPR (inegalităţile Bell, etc.; conform paragrafelor 5.3, 5.4, 5.17). Ne-am putea imagina, de exemplu, măsurători făcute în două puncte distincte ale unui microtubul pentru microtubule separate – care să arate că rezultatele măsurătorilor nu pot fi explicate în termenii unor acţiuni clasice independente care au loc în cele două puncte.

Oricare ar fi statutul unor asemenea sugestii, cercetarea microtubulelor este încă la început. Nu am nici o îndoială că acest domeniu ne mai rezervă încă surprize semnificative.

7.6. Microtubulele şi conştiinţa

Am putea fi întrebaţi dacă există vreo dovadă că fenomenul conştiinţei este legat în vreun fel cu acţiunea citoscheletului şi a microtubulelor? Ei bine, o astfel de dovadă există şi în cele ce urmează vom examina natura ei particulară care abordează problema conştiinţei din perspectiva cauzelor care îi produc absenţa!

Un drum important către răspunsul la întrebarea privind bazele fizice ale conştiinţei vine chiar din examinarea cauzelor care îi produc extincţia. *Anestezia generală* are exact această proprietate – complet reversibilă, dacă concentraţiile nu sunt prea mari. Este interesant faptul că anestezia generală poate fi indusă de un număr mare de substanţe complet diferite care par să nu aibă nici un fel de relaţie chimică una cu alta. Pe lista anesteticelor generale putem găsi substanţe atât de diferite ca oxidul de azot (N_2O), eterul ($CH_3CH_2OCH_2CH_3$), cloroformul($CHCl_3$), halotanul ($CF_3CHClBr$), izofloranul ($CHF_2OCHClCF_3$) şi chiar xenonul, gaz inert!

Dacă nu chimia este responsabilă pentru anestezia chimică, atunci ce anume o produce? Există şi alte tipuri de interacţiuni care au loc între molecule, intracţiuni mult mai slabe decât forţele chimice. Una dintre acestea se numeşte forţă Van der

Waals şi este o atracţie electrică slabă între moleculele care au *moment electric de dipol* (echivalentul electric al momentului magnetic de dipol care măsoară tăria magneţilor obişnuiţi). Amintiţi-vă că dimerii tubulinici sunt capabili să apară în două conformaţii diferite. Acestea par să apară din cauza unui electron plasat central în regiunea din fiecare dimer în care nu există apă. Electronul poate ocupa una din cele două poziţii separate. Forma dimerului este afectată de această poziţionare, adică de momentul electric de dipol. Abilitatea dimerului de a trece de la o conformaţie la alta este influenţată de forţa Van der Waals exercitată de către substanţele înconjurătoare (în regiunile 'hidrofobe' de unde apa a fost eliminată vezi Franks şi Liebb, (1982)) care interferează cu acţiunile normale ale tubulinelor. Pe măsură ce gazele anesteziante difuzează în celulele nervoase individuale, proprietăţile lor de dipol electric (care nu au mare lucru de-a face cu proprietăţile chimice) pot să întrerupă funcţionarea microtubulelor. Iată un mod cu totul plauzibil în care se poate petrece anestezia generalizată. Cu toate că nu pare să existe un consens general privind acţiunea detaliată a anesteziceior, o descriere coerentă este cea care atribuie rolul principal interacţiunilor Van der Waals dintre aceste substanţe şi dinamica comformaţională a proteinelor din creier. Este foarte posibil ca proteinele relevante să fie dimerii tubulinici din microtubulele neuronale şi deci să apară, în pierderea conştiinţei, întreruperea funcţionării microtubulelor.

Ca sprijin pentru sugestia conform căreia citoscheletul este cel direct afectat de anestezia generală, se poate remarca că nu doar 'animalele superioare', ca mamiferele sau păsările sunt anesteziate de aceste substanţe. Anesteziantele afectează în mod similar parameciul sau amoeba, în concentraţii asemănătoare cu cele folosite de medicina umană. Deşi nu ştim dacă cilii sau centriolul parameciului sunt afectaţi de anestezic, putem spune cu siguranţă că trebuie să fie o parte a citoscheletului său. Dacă acceptăm că sistemul de control al unui astfel de animal monocelular este într-adevăr citoscheletul, atunci afirmaţia că anesteziculi trebuie să lucreze asupra unei părţi a citoscheletului este consistentă cu restul descrierii.

Toate astea nu înseamnă că organismele monocelulare trebuie considerate conştiente. Dimpotrivă, problema conştiinţei este cu desăvârşire alta: ce anume *se adaugă* citoscheletelor pentru a 'trezi' o stare de conştiinţă. Pe de altă parte, argumentele discutate până acum par să indice că starea (stările) de conştiinţă are drept condiţii de posibilitate funcţionarea unui astfel de citoschelet. Fără un sistem de citoschelete în funcţiune conştiinţa este extinctă; în momentul în care

funcţionarea citoscheletelor este inhibată, conştiinţa dispare, revenind imediat ce acestea sunt puse în funcţiune, dacă pe parcurs nu s-a produs nici o defecţiune. Desigur că se poate pune întrebarea dacă nu cumva un parameci sau, la urma urmei, o celulă individuală din ficat, de exemplu ar putea poseda o anumită formă rudimentară de conştiinţă, însă această întrebare nu poate căpăta răspuns prin asemenea consideraţii. În afară de aceasta, orice descriere a conştiinţei trebuie să ia în considerare faptul că organizarea neuronală a creierului este fundamental implicată în guvernarea *formei* pe care o ia conştiinţa. Dacă această organizare a creierului nu ar fi esenţială, ficatul ar fi capabil să producă conştiinţă în aceeaşi măsură în care o face creierul.

Pe de altă parte, discuţiile de mai sus lămuresc că nu doar reţeaua neuronală a creierelor noastre contează în 'producerea' conştiinţei. Alcătuirile citoscheletice ale acestor neuroni par să fie importante în aceeaşi măsură.

Foarte probabil, pentru ca apariţia conştiinţei să se poată produce, în general, nu citoscheletul *ca atare* este cel relevant, ci o anumită *acţiune fizică esenţială* incorporată ingenios de biologie în activitatea microtubulelor. Dar ce anume poate fi ea? Esenţa argumentelor din partea întâia a fost că avem nevoie de ceva aflat dincolo de posibilităţile de simulare computaţională dacă vrem să găsim o bază pentru actele conştiente. Argumentaţia din partea a doua, în capitolele precedente, ne-a sugerat să privim la graniţa dintre nivelul cuantic şi nivelul clasic, acolo unde fizica contemporană ne obligă să folosim procedura de reducţie **R**, şi unde am susţinut de ce anume este nevoie de o nouă teorie fizică, **OR**. În acest capitol, am încercat să indic locul din creier în care acţiunea cuantică poate fi importantă pentru comportamentul clasic, iar argumentele ne-au dus la concluzia că interfaţa cuantic/clasic se manifestă prin *controlul citoscheletic al conexiunilor sinaptice*. Să explorăm acum această descriere mai în detaliu.

7.7. Un model pentru minte?

Aşa cum am remarcat în paragraful 7.1, pare potrivit să acceptăm faptul că semnalele nervoase sunt, în sine, lucruri care pot fi tratate într-un mod strict clasic, pe motivul că este foarte probabil ca astfel de semnale să producă în mediu perturbaţii la un nivel care face imposibilă menţinerea coerenţei cuantice. Dacă

conexiunile sinaptice şi tăria lor sunt considerate fixate, atunci modul în care fiecare neuron emite şi îl afectează pe următorul este din nou un lucru care poate fi tratat clasic, cu excepţia elementului aleator care apare în cursul acestui proces. Acţiunea creierului în astfel de circumstanţe ar fi complet calculabilă, în sensul că o simulare computaţională ar fi, în principiu, posibilă. Prin asta nu înţeleg că o astfel de simulare ar imita în mod precis acţiunile unui creier *specific* din cauza elementelor aleatoare – ci ar fi o simulare a unei acţiuni *tipice* a creierului şi deci un comportament tipic al unui individ controlat prin intermediul unui astfel de creier (vezi paragraful 1.7). Mai mult, această afirmaţie este o afirmaţie de principiu şi nu implică în nici un fel faptul că o astfel de simulare ar putea fi chiar realizată cu tehnologia zilelor noastre. Mai trebuie de asemenea presupus că ingredientele aleatoare sunt *într-adevăr* aleatoare. Posibilitatea unei 'minţi externe', dualiste, care să influenţeze aceste probabilităţi nu este luată în considerare în nici un fel (vezi paragraful 7.1).

Acceptăm astfel, cel puţin în mod provizoriu, că dacă avem conexiuni sinaptice *fixe*, creierul se comportă într-adevăr ca un fel de *calculator* – dar un calculator cu ingrediente aleatoare încorporate în el. Aşa cum am văzut din dicuţiile din partea întâia, este foarte improbabil ca o astfel de schemă să poată vreodată furniza un model pentru înţelegerea umană conştientă. Pe de altă parte, dacă conexiunile sinaptice specifice care definesc calculatorul neuronal particular sunt subiectul unei continue schimbări, iar controlul acestor schimbări este guvernat de o anumită acţiune necalculabilă, atunci ar fi într-adevăr posibil ca un astfel de model extins să simuleze comportamentul unui creier conştient.

Care poate fi, însă, această acţiune necalculabilă? În toată argumentaţia noastră trebuie să avem în vedere natura *globală* a conştiinţei. Dacă am avea pur şi simplu cazul a 10^{11} citoschelete individuale care contribuie fiecare la un *input* necalculabil, ar fi greu de văzut în ce măsură ne-ar fi în vreun fel de ajutor. Conform argumentelor din partea întâia, comportamentul necalculabil este legat într-adevăr de acţiunea conştientă cel puţin în măsura în care *o anumită* acţiune conştientă şi în special *înţelegerea* par să fie necalculabile. Însă acest lucru nu este relevant pentru citoscheletele individuale sau pentru microtubulele individuale din interiorul unui citoschelet. Nu poate fi vorba despre un citoschelet sau microtubul particular care să 'înţeleagă' vreo parte din teorema lui Gödel! Înţelegerea este ceva care operează la o scară globală; iar dacă citoscheletele sunt implicate, atunci

trebuie să fie vorba de un fenomen colectiv care cuprinde un număr mare de citoschelete în acelaşi timp.

Amintiţi-vă ideea lui Fröhlich conform căreia fenomene cuantice la scară mare -- poate de tipul condensării Bose-Einstein -- ar putea fi o posibilitate biologică definită, chiar în interiorul creierului 'fierbinte' (Marshall, 1989). Ceea ce avem în vedere aici este că într-o astfel de stare ar trebui să fie implicate nu doar microtubulele individuale; o astfel de stare ar trebui să se extindă de la un microtubul la următorul. Astfel, nu numai coerenţa cuantică trebuie să se întindă pe lungimea unui microtubul întreg (şi să ne amintim că microtubulele pot fi destul de lungi), ci marea majoritate a microtubulelelor citoscheletului unui neuron, dacă nu toate, trebuie să fie în aceeaşi stare cuantică. Şi nu numai atât: coerenţa cuantică trebuie să 'sară' bariera sinaptică dintre neuroni, pentru a se extinde la neuronul următor. De altfel n-ar putea fi vorba de globalitate, dacă ea ar cuprinde o singură celulă... Unitatea unei minţi poate apare, într-o astfel de descriere, doar dacă există o formă de coerenţă cuantică care să se extindă de-a lungul unei părţi apreciabile a întregului creier.

Ori, o astfel de proprietate ar fi absolut remarcabilă aproape incredibilă – dacă ar fi realizată cu mijloace biologice. Şi totuşi, cred că avem bune motive să bănuim că Natura a reuşit chiar acest lucru; iar dovada principală ne este dată de chiar existenţa mentalului. Mai sunt multe lucruri încă neînţelese despre sistemele biologice şi felul magic în care funcţionează. Există multe lucruri în biologie care depăşesc cu mult tot ceea ce putem realiza în prezent cu mijloacele noastre tehnice (gândiţi-vă, de exemplu, la un păianjen de câţiva milimetri care îşi ţese delicata şi complicata lui pânză). Să ne amintim că anumite efecte de coerenţă cuantică la distanţă de câţiva metri au fost deja observate (cu metode *fizice*) experimental (vezi paragraful 5.4). În ciuda dificultăţii tehnice pe care astfel de experimente o presupun, n-ar trebui să excludem posibilitatea ca Natura să fi găsit căi biologice de a realiza mult mai mult. 'Ingenuitatea' pe care o găsim mereu în sistemele biologice nu trebuie niciodată subestimată.

Cu toate acestea, argumentaţia pe care am prezentat-o cere mult mai mult decât coerenţă cuantică la distanţă mare. Ea cere, în primul rând, ca sistemele biologice care sunt creierele noastre să fi ajuns cumva la manevrarea detaliilor unei fizici încă necunoscută fizicienilor. Această fizică este teoria lipsă OR care uneşte nivelul clasic şi cel cuantic şi, aşa cum argumentam, înlocuieşte procedura de

reducţie **R** printr-o schemă fizică subtilă, necalculabilă (dar fără îndoială tot matematică).

Faptul că fizicienii n-au reuşit să afle până acum care este această 'teorie lipsă' nu este, desigur, un argument împotriva presupoziţiei că Natura foloseşte rezultatele ei în biologie. Ea a folosit principiile mecanicii newtoniene mult înaintea lui Newton, fenomenele electromagnetice înaintea lui Maxwell, iar mecanica cuantică înaintea lui Plank, Einstein, Bohr, Heisenberg, Schroedinger şi Dirac — ba chiar cu câteva milioane bune de ani mai înainte! Doar aroganţa timpului nostru i-a putut face pe unii să creadă că ştim toate principiile fundamentale care stau la baza tuturor subtilităţilor din biologie. Atunci când un organism este destul de norocos să dea peste o asemenea acţiune subtilă, evident că va profita de beneficiile pe care un proces fizic de acest fel i le-ar putea aduce. Iar Natura îi surâde, lui şi descendenţilor lui, şi permite acţiunii fizice în chestiune să se păstreze de la o generaţie la alta, multiplicându-şi ocurenţele, prin puternicul proces de selecţie naturală.

Atunci când au apărut primele celule eucariote, ele trebuie să fi găsit un mare câştig în prezenţa microtubulelor primitive în interiorul lor. Apoi a apărut, probabil, un anumit tip de influenţă organizaţională, ca aceea pe care am prezentat-o şi care le-a permis să se comporte într-un mod încă rudimentar dar orientat către un scop, iar acest lucru le-a permis să supravieţuiască mai bine decât competitorii lor. Ar fi, fără îndoială, nepotrivit să ne referim la o asemenea influenţă ca la o 'minte'; şi totuşi ea a apărut, cred, la interfaţa dintre procesele de nivel clasic şi cele de nivel cuantic. Mai mult, natura subtilă a acestei interacţiuni este determinată de acţiunea fizică sofisticată a procesului **OR** — încă necunoscut în detaliu — care, în circumstanţe mai puţin organizate, se manifestă sub forma crudă a procesului cuantic de reducţie **R**. Descendenţii îndepărtaţi ai acestor organisme monocelulare — parameciul de astăzi, amoeba, dar şi furnicile, copacii, broaştele, floarea de untişor sau fiinţele umane — toate au păstrat beneficiile pe care această acţiune sofisticată le-a adus primelor creaturi celulare pe care le-a modificat substanţial, pentru a servi unor scopuri complet diferite. Însă doar atunci când a fost încorporată într-un sistem nervos foarte dezvoltat a fost aceasta în stare să-şi realizeze într-o bună măsură imensul său potenţial — şi să realizeze ceea ce astăzi numim 'minte'.

Să acceptăm deci posibilitatea ca totalitatea microtubulelor din citoscheletele

unei largi familii de neuroni din creierele noastre să ia parte într-un fenomen de coerenţă cuantică globală — sau, cel puţin, să presupunem că există suficientă nonlocalizare cuantică între stările diferitelor microtubule din creier — astfel încât o descriere *clasică* globală a acţiunii lor colective să *nu fie* corespunzătoare. Am putea să luăm în considerare complicate 'oscilaţii cuantice' în interiorul microtubulelor, în care stratul izolator al tuburilor este suficient pentru a ne asigura că nu se pierde întreaga coerenţă cuantică. Foarte tentantă este presupoziţia conform căreia putem pune în legătură calculele de tip automate celulare despre care se presupune că au loc *de-a lungul tuburilor* (Hameroff şi colegii săi), cu presupusele oscilaţii cuantice (de exemplu cele propuse de del Giudice et.al, 1983, sau de Jibu et.al, 1994) care au loc *în interiorul tuburilor*.

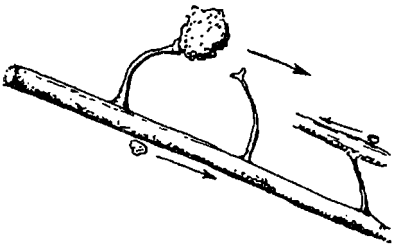


Fig. 7.13. MAP transportă molecule mari, pe când alte molecule se mişcă direct de-a lungul microtubulelor.

În aceeaşi ordine de idei, putem remarca că tipul de frecvenţă sugerat de Fröhlich pentru oscilaţiile cuantice 'colective şi confirmat de observaţiile lui Grundler şi Keilmann (1983) — în regiunea de frecvenţă de 5×10^{10} Hz (adică 5×10^{10} oscilaţii pe secundă) — este acelaşi cu tipul de frecvenţă considerat, de Hameroff şi colegii săi ca 'timp de întrerupere' pentru dimerii tubulinici în automatele celulare microtubulare. Astfel, dacă mecanismul lui Fröhlich este într-adevăr cel care operează în microtubule, atunci asta indică un anumit tip de cuplaj între cele două tipuri de activitate[7]

[7] Este mult mai puţin clar, totuşi, dacă poate exista vreo conexiune directă între astfel de frecvenţe ridicate şi mult mai familiara activitate a undelor creierului (ca, de exemplu, ritmul alfa, la 8–12 Hz). Se poate imagina că aceste frecvenţe joase pot apărea ca frecvenţe de

Totuşi, dacă cuplajul este prea puternic, nu este posibilă menţinerea naturii cuantice pentru oscilaţiile interne fără a pretinde în acelaşi timp calculelor efectuate de-a lungul tuburilor să fie tratate cu ajutorul mecanicii cuantice. Iar dacă acesta este cazul, atunci vom avea, de-a lungul microtubulelor, un anumit tip de *calcul cuantic*! (vezi paragraful 7.3) Trebuie să ne întrebăm, deci, dacă această posibilitate este serioasă. Dificultatea principală este că ea ar părea să ceară ca schimbările în conformaţiile dimerilor să nu perturbe semnificativ materia aflată împrejur. Trebuie să subliniem, în legătură cu acest lucru, că regiunea aflată împrejurul microtubulelor pare să conţină apă într-o structură ordonată, în timp ce alte substanţe sunt excluse (vezi Hameroff, 1987, p. 172), ceea ce ne poate da o măsură a ecranării cuantice. Pe de altă parte, există MAP (proteine asociate microtubulelor) de conexiune (vezi paragraful 7.4) care ies din microtubule, unele cu rolul de a transporta alte substanţe, şi părând a fi influenţate de mişcarea semnalelor de-a lungul tuburilor (vezi Hameroff, p.122). Acest din urmă fapt pare să ne spună că preocesele de calcul care se petrec în tub pot într-adevăr perturba mediul într-un grad suficient încât să poată fi tratate clasic. Gradul de perturbare rămâne destul de mic în termenii mişcărilor de masă, conform criteriului OR propus în paragraful 6.12, însă pentru ca sistemul întreg să rămână la nivel cuantic, ar fi necesar ca aceste perturbaţii să nu se extindă în celulă, sau dincolo de graniţele acesteia. După părerea mea, există un grad suficient de incertitudine, atât în relaţie cu situaţia fizică actuală cât şi cu felul în care criteriul OR urmează să fie aplicat, pentru ca să putem fi siguri că o imagine, în întregime clasică, este sau nu adecvată în acest stadiu.

Să presupunem, totuşi, de dragul discuţiei, că am putea trata clasic calculele care au loc în microtubule, în sensul că nu vom considera relevante superpoziţiile cuantice ale diferitelor calcule. Pe de altă parte, să presupunem de asemenea că există oscilaţii cuantice reale, de un tip oarecare, în interiorul tuburilor, cu un anumit tip de cuplaj fin între aspectul cuantic intern şi cel extern, clasic, al fiecărui tub. Conform acestei imagini, detaliile teoriei noi OR, de care avem nevoie, vor juca un rol cu adevărat important pentru acest cuplaj: va trebui

---

rezonanţă, însă nu s-a putut stabili nici o legătură între ele. Sunt interesante în legătură cu acestea, observaţiile recente privind oscilaţiile de 35–75 Hz care par să apară în asociaţie cu regiunile din creier care sunt implicate în fenomenul de atenţie conştientă. Aceste oscilaţii par să aibă anumite efecte non–locale uimitoare (vezi Eckhorn et.al., 1988, Graz şi Singer, 1989, Crick şi Koch, 1990, 1992, Crick, 1994).

să luăm în considerare o anumită influenţă a oscilaţiilor cuantice interioare asupra calculelor exterioare, lucru care pare rezonabil dacă luăm în considerare mecanismele pe care le-am presupus responsabile pentru comportamentul de automate celulare al microtubulelor, şi anume influenţele de tip Van der Waals dintre dimerii tubulinici învecinaţi.

Imaginea pe care o construim va fi deci cea a unei stări cuantice globale care cuplează în mod coerent activităţile care au loc în interiorul tuburilor, pentru colective de microtubule de-a lungul unor arii considerabile din creier. Există o anume influenţă pe care această stare (care ar putea să nu fie doar o simplă 'stare cuantică', în sensul convenţional al formalismului cuantic standard) o exercită asupra calculelor care au loc de-a lungul microtubulelor o influenţă determinată de fizica non-computaţională OR, încă necunoscută, în favoarea căreia am pledat până acum. Activitatea 'computaţională' de schimbare comformaţională în tubuline controlează felul în care tuburile transportă materiale de-a lungul suprafeţei lor exterioare (vezi figura 7.13) şi, în ultimă instanţă, influenţează tăriile sinapselor la terminaţiile pre şi post-sinaptice. În acest fel, o mică parte a acestei organizări cuantice coerente din interiorul microtubulelor este 'interceptată' pentru a influenţa schimbările în conexiunile sinaptice ale calculatorului neuronal.

O astfel de imagine este deschisă tuturor speculaţiilor. De exemplu, putem găsi un posibil rol pentru enigmatica non-localitate de tip EPR, sau pentru straniile contrafactuale din mecanica cuantică. Calculatorul neuronal este, poate, echilibrat să efectueze anumite calcule pe care nu le calculează în mod actual, însă (ca în cazul problemei cu testarea bombelor), simplul fapt că *le-ar fi putut* face, produce un efect diferit de cel corespunzător cazului în care i-ar fi fost imposibil să le efectueze. În acest fel, 'reţeaua' clasică a calculatorului neuronal în fiecare moment, ar putea avea o influenţă asupra stării interne a citoscheletului, chiar dacă emisiile neuronice care ar activa în mod normal acea configuraţie a calculatorului ar putea, de fapt, să nu aibă loc. Am putea găsi analogii posibile cu această imagine în multe dintre activităţile mentale familiare în care suntem continuu implicaţi — dar simt că este mai bine să nu continui cu astfel de lucruri.

În imaginea pe care o sugerez ca ipoteză, conştiinţa ar fi o anumită manifestare a acestei stări cuantice non-locale a citoscheletelor în relaţie (tip OR) cu mediul lor, la interfaţa dintre nivelul cuantic şi nivelul clasic de activitate. Sistemul de neuroni interconectaţi, asemănător unui calculator, va fi influenţat în

mod continuu de către această activitate citoscheletică, ca manifestare a acelui fenomen (oricare ar fi el) pe care-l numim *liber arbitru*. Rolul neuronilor în această imagine este mai degrabă unul de aparat de mărire, în care acţiunea citoscheletică la scară mică este transferată în ceva care poate influenţa alte organe ale corpului – ca, de exemplu, în muşchi. În mod corespunzător, nivelul neuronal al descrierii care furnizează imaginea curentă la modă a creierului şi a minţii, este doar o *umbră* a nivelului mai adânc al acţiunii citoscheletului. Şi acolo trebuie să căutăm bazele fizice ale *minţii*.

Desigur că există o cantitate mare de speculaţie în această imagine; ea nu este însă în afara zonei de acţiune a înţelegerii ştiinţifice curente. Am văzut în capitolul precedent că există motive puternice pentru a aştepta ca ideile fizicii curente să fie modificate; motive provenind, în mare parte, chiar din consideraţii ale fizicii cuantice asupra unor fenomene care au loc la o scară relevantă pentru funcţionarea microtubulelor sau pentru interfaţa dintre citoschelet şi neuron. Din argumentele prezentate în partea întâia am tras concluzia că avem nevoie de o deschidere spre o acţiune fizică non-computaţională dacă este să găsim un 'loc' pentru conştiinţă. În partea a doua am arătat că unicul loc plauzibil pentru această acţiune este dat de înlocuirea procesului de reducţie a stării cuantice, **R**, cu mecanismul **OR**. Acum trebuie să vedem dacă există vreun temei pur fizic pentru convingerea noastră că **OR** poate avea într-adevăr o natură non-computaţională. Vom vedea că, urmând sugestiile din paragraful 6.12, vom găsi astfel de temeiuri.

7.8. Non-computabilitate în gravitaţia cuantică (1)

O cerinţă cheie în discuţia precedentă era ca, oricare ar fi 'noua fizică' ce vine să înlocuiască procedura probabilistă **R**, ea să aibă un aspect non-computaţional. Am susţinut în paragraful 6.10 faptul că această nouă fizică, **OR** trebuie să combine principiile teoriei cuantice cu cele ale relativităţii generalizate: trebuie să fie, deci un fenomen de *gravitaţie cuantică*. Există oare vreun fundament pentru presupunerea că necalculabilitatea ar fi o trăsătură esenţială a oricărei teorii ce ar putea apare şi unifica teoria cuantică şi gravitaţia generalizată?

Într-o abordare particulară a gravitaţiei cuantice, Robert Geroch şi James Hartle (1986) au ajuns la o problemă insolubilă în mod computaţional, *problema*

de echivalenţă topologică pentru varietăţile 4-dimensionale. În linii mari, problema lor este de a decide când anume două spaţii 4-dimensionale sunt 'identice' din punct de vedere topologic (adică, în ce condiţii este posibilă deformarea continuă a unuia dintre ele până când ajunge să coincidă cu celălalt; nu sunt permise operaţiile de 'tăiere' sau lipire asupra celor două spaţii). În figura 7.14 această problemă este ilustrată pe cazul 2-dimensional: suprafaţa unei ceşti este topologic aceeaşi cu suprafaţa unui inel, însă diferită de suprafaţa unei mingi. În două dimensiuni, problema echivalenţei topologice se poate rezolva computaţional; în cazul patru-dimensional, însă, nu există un algoritm care să o poată rezolva (după cum a arătat A.A Markov (1958)). Demonstraţia arată, de fapt, că dacă ar exista un astfel de algoritm, atunci el ar putea fi convertit într-unul care să rezolve *problema opririi*, adică, să decidă dacă o maşină Turing se va opri sau nu. Dar, aşa cum am văzut în paragraful 2.5, nu există un asemenea algoritm şi deci rezultă că nu poate exista nici un algoritm care să rezolve problema echivalenţei pentru varietăţi patru-dimensionale.

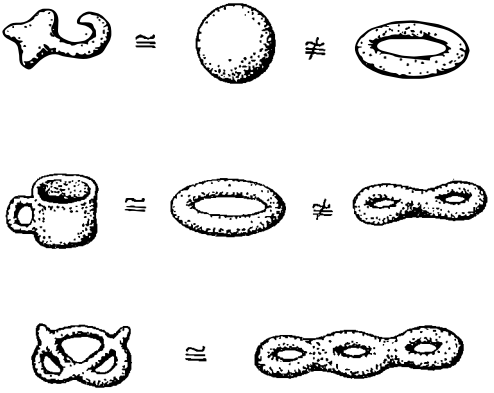


Fig. 7.14. Suprafeţele bidimensionale închise pot fi clasificate computaţional (de ex. prin numărul de 'mânere'). Dar 'suprafeţele' patru-dimensionale închise nu pot fi clasificate computaţional.

Există multe alte clase de probleme matematice care sunt nerezolvabile computaţional. Două dintre ele, a zecea problemă a lui Hilbert şi problema acoperirii, au fost discutate în paragraful 1.9 (pentru un alt exemplu, vezi **MN** p. 43, 67-74).

Trebuie să subliniez că „imposibil de rezolvat computaţional" nu înseamnă că există vreo problemă din clasa citată care să fie insolubilă în principiu, în mod individual. Ceea ce susţine afirmaţia precedentă este că nu există nici o metodă sistematică (algoritmică) de a rezolva toate problemele din clasa respectivă. În unele din cazurile individuale putem ajunge la o soluţie cu ajutorul ingeniozităţii şi intuiţiei, ajutate eventual de calcul, aşa cum pot exista cazuri în care s-ar putea ca soluţia să fie inaccesibilă minţii umane. Nu există o soluţie generală a problemei, care este astfel deschisă interpretărilor. Totuşi, ceea ce arată discuţia din paragraful 2.5, împreună cu argumentele din capitolul 3, este că aceste probleme, chiar dacă individual accesibile unei rezolvări intuitive, formează o clasă inaccesibilă computaţional (în cazul problemei de oprire, paragraful 2.5, rezultă că mulţimea de calcule despre care se poate arăta că nu se opresc nu poate fi cuprinsă în nici un algoritm corect A, iar argumentele din capitolul 3 continuă din acest punct).

În abordarea Geroch-Hartle a mecanicii cuantice intră în discuţie problema de echivalenţă a varietăţilor patru-dimensionale deoarece, conform regulilor standard ale teoriei cuantice, starea cuantic-gravitaţională va cuprinde superpoziţii ale tuturor geometriilor posibile aici geometrii spaţio-temporale, adică obiecte patru-dimensionale, cu factori de pondere numere complexe. Pentru a înţelege cum pot fi specificate astfel de superpoziţii într-un mod unic (adică fără a 'repeta numărătoarea') este necesar să ştim când putem considera că astfel de spaţii-timp sunt diferite şi când pot fi considerate identice. Problema echivalenţei topologice apare deci ca o parte a acestei decizii.

S-ar putea pune întrebarea: dacă ceva din natura teoriei Geroch-Hartle se dovedeşte corect fizic, înseamnă asta că există o parte esenţial necalculabilă în evoluţia unui sistem fizic? Nu cred că răspunsul la această întrebare este prea clar. Nu-mi este clar nici măcar dacă imposibilitatea de a rezolva echivalenţa topologică atrage după sine imposibilitatea de a rezolva problema generală a echivalenţei *geometrice*. Este de asemenea discutabil cum (şi dacă) poate fi legată această abordare de ideile OR pe care le-am sugerat, unde trebuie să ne aşteptăm la o schimbare actuală în structura teoriei cuantice exact în momentul în care efectele gravitaţionale devin relevante.

Cu toate acestea, teoria Geroch-Hartle ne arată posibilitatea clară ca necalculabilitatea să aibă un rol real într-o teorie a gravitaţiei cuantice, indiferent

care va fi forma ei finală.

7.9. Maşini oracol şi legi fizice

Lăsând la o parte toate cele de mai înainte, ar fi timpul să punem o întrebare separată. Să presupunem că teoria emergentă a gravitaţiei cuantice va fi într-adevăr non-computaţională, în sensul că va permite construirea unui dispozitiv fizic care va rezolva problema de oprire. Ar fi acest lucru suficient pentru a rezolva toate problemele care apar din consideraţiile noastre asupra argumentelor Gödel-Turing din partea întâia? În mod poate surprinzător, răspunsul la această întrebare este 'nu'.

Să vedem de ce posibilitatea de rezolvare a problemei de oprire nu ne ajută. În 1939, Turing a introdus un concept relevant pentru această problemă, pe care l-a denumit *oracol*. Un oracol ar fi un obiect (fictiv, neconstruibil fizic, existent doar mental) capabil să rezolve problema de oprire. Astfel, dacă îi arătăm oracolului perechea de numere naturale q,n, el ne va da, după un timp finit, răspunsurile **DA** sau **NU**, în cazurile în care calculul $C_q(n)$ ajunge sau nu ajunge la final (vezi paragraful 2.5). Argumentele din 2.5 ne furnizează o demonstraţie a rezultatului obţinut de Turing, conform căruia nu se poate construi un astfel de oracol care să lucreze într-un mod complet calculabil, însă nu ne spune că un oracol n-ar putea fi construit din punct de vedere fizic. Pentru ca această concluzie să rezulte, am avea nevoie să ştim că legile fizicii sunt calculabile prin natura lor ceea ce este chiar subiectul Părţii a doua a acestei cărţi. Trebuie, de asemenea, subliniat că posibilitatea fizică de a construi un astfel de oracol nu are, din câte pot spune, o implicaţie asupra punctului de vedere pe care îl susţin. Aşa cum am menţionat anterior, nu există o cerinţă ca toate problemele de oprire să fie accesibile intelectului sau intuiţiei umane, astfel că nu există nici un motiv să presupunem că ele ar fi la îndemâna vreunui aparat construibil.

Turing a luat în considerare în discuţia acestor probleme o modificare a calculabilităţii astfel încât să putem introduce un astfel de oracol în orice stadiu dorim. Astfel, o *maşină oracol* (care acţionează ca un *algoritm oracol*) ar fi o maşină Turing obişnuită, cu excepţia faptului că, în plus faţă de operaţiile normale,

calculabile, este în stare să efectueze o altă operaţie, descriptibilă după cum urmează: „Cheamă oracolul şi întreabă-l dacă $C_q(n)$ se opreşte; dacă obţii răspunsul, continuă calculul şi foloseşte răspunsul obţinut" Oracolul poate fi întrebat din nou şi din nou, dacă este nevoie. Observaţi că o maşină oracol este tot atât de deterministă ca şi o maşină Turing obişnuită, ceea ce ilustrează faptul că determinismul şi calculabilitatea nu sunt acelaşi lucru. În principiu, este tot atât de posibil să avem un univers care se comportă determinist ca o maşină oracol, ca şi unul care evoluează determinist ca o maşină Turing (universurile de jucărie descrise în paragraful 1.9 şi la p. 187 din **MN** sunt, de fapt, universuri maşini-oracol.)

Evoluează deci universul nostru ca o maşină oracol? Este interesant că argumentele din partea întâia a cărţii se pot aplica în egală măsură şi aproape fără modificări împotriva unui model al înţelegerii matematice de tipul maşină-oracol ca şi împotriva modelului maşină-Turing. Tot ce trebuie să facem, în discuţia din 2.5, este să citim '$C_q(n)$' ca fiind „cea de-a q-a maşină oracol aplicată numărului natural n " Să rescriem asta ca $C'_q(n)$. Maşinile oracol pot fi listate computaţional la fel ca maşinile Turing obişnuite. Singura trăsătură adiţională pe care trebuie s-o luăm în considerare este că trebuie să observăm stadiile de calcul în care oracolul este întrebat lucru care, de altfel, nu prezintă nici o dificultate. Vom înlocui acum *algoritmul* $A(q,n)$ din paragraful 2.5 cu un algoritm oracol $A'(q,n)$ despre care vom considera că reprezintă totalitatea mijloacelor aflate la îndemâna intelectului uman pentru a decide cu certitudine că operaţia $C'_q(n)$ nu se sfârşeşte. Refăcând argumentaţia exact ca înainte, tragem concluzia:

$\mathcal{G}'$ Matematicienii nu folosesc un algoritm-oracol corect cognoscibil pentru a stabili adevărul matematic.

Din aceasta, tragem concluzia că nici o fizică care lucrează ca o maşină oracol nu ne va rezolva problemele. De fapt, întregul proces poate fi repetat şi aplicat unor 'maşini oracol de ordinul doi' care ne pot spune dacă o maşină oracol obişnuită se opreşte vreodată. Ca şi mai înainte, ajungem la concluzia:

$\mathcal{G}''$ Matematicienii nu folosesc un algoritm-oracol de ordinul doi corect cognoscibil pentru a stabili adevărul matematic.

Ar trebui să fie clar că procesul se poate repeta din nou şi din nou, în felul gödelizării repetate, aşa cum a fost discutat în răspunsul la **Q19**. Pentru orice ordinal recursiv calculabil α avem un concept de maşină oracol de ordinul α şi

putem concluziona că:

$\mathcal{G}^{\alpha}$ Matematicienii nu folosesc maşini-oracol de ordinul α corecte cognoscibile pentru a stabili adevărul matematic, pentru orice ordinal calculabil α.

Concluzia finală pentru toate acestea este mai degrabă alarmantă, pentru că sugerează că trebuie să căutăm o teorie fizică non-computaţională care trece dincolo de nivelul oricărei maşini oracol.

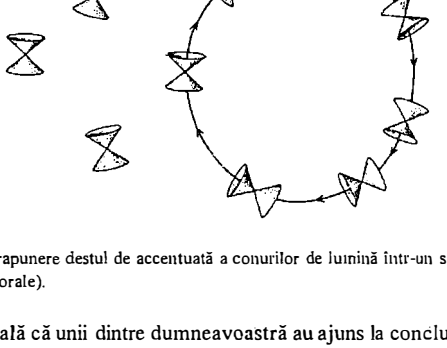


Fig. 7.15. La o suprapunere destul de accentuată a conurilor de lumină într-un spaţiu-timp, pot apare curbe închise (temporale).

Fără îndoială că unii dintre dumneavoastră au ajuns la concluzia că am depăşit limita credibilităţii cu această discuţie! Cu siguranţă, nu pot blama pe nici un cititor care simte astfel. Însă aceste sentimente nu sunt o scuză suficientă pentru a nu lua în considerare în detaliu argumentele pe care vi le ofer. În particular, argumentele din capitolele 2 şi 3 trebuie toate recitite, înlocuind maşinile Turing obişnuite cu maşini oracol de ordin α. Nu cred că discuţia va fi afectată în vreun fel, însă trebuie să mărturisesc că nu îndrăznesc să vă plictisesc cu ea din nou. Trebuie totuşi subliniat că nu este necesar ca înţelegerea umană a matematicii să fie la fel de puternică (în principiu), ca orice maşină oracol. Aşa cum am mai observat, concluzia $\mathcal{G}$nu implică în mod *necesar* că intuiţia umană este suficient de puternică, teoretic, pentru a rezolva orice problemă de oprire. Astfel, nu avem nevoie să concluzionăm că legile fizice pe care le căutăm trebuie să depăşească, în principiu, orice nivel calculabil al unei maşini oracol (sau chiar să egaleze ordinul întâi al unei astfel de maşini). Trebuie să căutăm doar ceva care nu este echivalent cu nici una

dintre maşinile oracol specifice (inclusiv maşinile de ordin zero, adică maşinile Turing). Legile fizice pot conduce la ceva care este doar *diferit*.

7.10. Non-computabilitate în gravitaţia cuantică (2)

Să ne întoarcem acum la problema gravitaţiei cuantice. Trebuie să subliniem că nu există, deocamdată, o teorie acceptată; nu există nici măcar un candidat acceptabil. Cu toate acestea, pot fi enumerate mai multe propuneri diferite şi fascinante [8]. Cea la care doresc să mă refer în mod special în cele ce urmează are în comun cu abordarea Geroch-Hartle cerinţa de a considera diferite superpoziţii cuantice spaţii-timp distincte (multe dintre abordări cer doar superpoziţii de geometrii spaţiale tridimensionale, ceea ce este oarecum diferit). Sugestia, datorată lui David Deutsch [9], este că putem suprapune, de-a lungul unor geometrii spaţio-temporale 'rezonabile' în care timpul se comportă normal, geometrii 'nerezonabile', în care există *linii de tip temporal închise*. Un astfel de spaţiu-timp este reprezentat în figura 7.15. O linie de tip temporal descrie posibilitatea istoriei unei particule clasice, 'tipul temporal' referindu-se la faptul că linia este întotdeauna direcţionată în interiorul conului luminos, în fiecare din punctele sale, astfel încât viteza absolută locală nu este depăşită — aşa cum cere teoria relativităţii (vezi 4.4). Semnificaţia unei linii de tip temporal închise este că putem contempla un 'observator' care are această linie ca linie de univers proprie[8], adică ca linie care descrie, în spaţiu-timp, istoria propriului său corp. Un astfel de observator se va găsi, după un parcurs finit în timpul pe care îl percepe în mod normal, înapoi în trecut (călătorie în timp, deci). Pare deschisă posibilitatea ca el să se afecteze cumva pe el însuşi, cel din trecut (presupunând că are un anumit fel de *liber arbitru*) astfel încât să producă evenimente şi transformări trecute pe care nu le-a experimentat niciodată, ceea ce duce la o contradicţie (de obicei, aceste discuţii îl pun să-şi omoare bunicul înainte ca tatăl lui să se fi născut sau altceva de acest tip, la fel de îngrijorător).

[8] Vezi nota către cititor de la începutul cărţii.

Argumente de acest fel ne oferă destule motive pentru a nu lua în serios spaţii-timp cu linii de tip temporal închise ca modele posibile ale universului actual. Destul de curios, Kurt Gödel a propus, în 1949, primele modele spaţio-temporale de acest tip. Gödel nu considera aspectul lor paradoxal ca un motiv adecvat pentru a le elimina ca modele cosmologice.

Pentru motive din cele mai diferite, suntem tentaţi astăzi să adoptăm o poziţie diferită, dar vezi, de exemplu, Thorne, (1994). Ar fi interesant să ne imaginăm ce reacţie ar avea Gödel în faţa folosirii unor astfel de spaţii-timp despre care va fi vorba în cele ce urmează!) Cu toate că pare într-adevăr rezonabilă eliminarea geometriilor spaţio-temporale cu linii de univers de tip temporal închise ca descrieri posibile ale universului *clasic*, se poate argumenta convingător faptul că ele nu trebuie neglijate ca ocurenţe potenţiale care ar putea fi implicate într-o superpoziţie cuantică. Exact aceasta este poziţia lui Deutsch. Deşi contribuţiile unor asemenea geometrii la vectorul de stare total pot fi foarte mici, prezenţa lor potenţială are, după Deutsch, un efect uluitor. Dacă considerăm acum ce înseamnă efectuarea unui calcul cuantic într-o astfel de situaţie, ajungem la concluzia că operaţii noncomputaţionale ar putea exista! Acest lucru apare din faptul că în geometriile spaţio-temporale cu linii de univers temporale închise, o maşină Turing ar putea să funcţioneze în cerc, indefinit, reluând propriile date de ieşire ca date de intrare, astfel încât răspunsul la întrebarea 'se opreşte calculul vreodată' să aibă o influenţă actuală asupra rezultatului final al calculului cuantic. Deutsch a ajuns la concluzia că în descrierea pe care el o dă problemei gravitaţiei cuantice, maşini oracol cuantice sunt posibile. În măsura în care am putut să-mi dau seama, argumentele sale se aplică în egală măsură maşinilor oracol de ordin superior.

Desigur că mulţi dintre dumneavoastră sunt deja de părere că toate acestea trebuie luate sub semnul îndoielii. Într-adevăr, nu există nici un argument hotărâtor care să facă din schema lui Deutsch o teorie a gravitaţiei cuantice, măcar plauzibilă dacă nu consistentă. Cu toate acestea, ideile avansate sunt coerente în contextul în care au fost avansate şi sunt interesante într-un mod sugestiv mi se pare deci rezonabil ca atunci când vom găsi o teorie corespunzătoare pentru gravitaţia cuantică, anumite elemente din propunerea lui Deutsch să supravieţuiască şi în aceasta din urmă. După părerea mea, o astfel de teorie va afecta chiar legile fundamentale ale mecanicii cuantice (vezi paragrafele 6.10, 6.12). Pe de altă parte, faptul că în abordarea lui Deutsch, non-computabilitatea este o trăsătură asociată

gravitaţiei cuantice este un sprijin considerabil pentru ideile mele privind posibilitatea unei acţiuni necalculabile.

În sfârşit, trebuie să observăm că tocmai deformarea potenţială a conurilor luminoase din relativitatea generalizată produce efectele noncomputaţionale indicate de Deutsch (vezi paragraful 4.4). Odată ce conurilor luminoase li se permite deformarea, chiar în măsura infimă care apare în teoria lui Einstein în circumstanţe obişnuite, apare *potenţial* posibilitatea unei deformări de o asemenea factură încât să producă linii temporale închise. Ori, în acord cu teoria cuantică, această posibilitate trebuie să joace un rol de tip contrafactual pentru ca să aibă efecte *actuale*!

7.11. Timpul şi percepţiile conştiente

Să ne întoarcem la problema conştiinţei. Rolul particular jucat de conştiinţă în percepţia adevărului matematic este ceea ce ne-a dus, de-a lungul acestui drum straniu, în mijlocul teritoriului străin în care ne aflăm acum. Evident, în conştiinţă există multe alte lucruri în afara percepţiei matematicii. Am urmat acest drum particular doar pentru că părea posibil să înaintăm de-a lungul lui. Fără îndoială că unii cititori nu se dau în vânt pentru acest 'undeva' unde ne găsim în momentul de faţă. Şi totuşi, dacă privim înapoi din punctul acesta, am putea descoperi că o parte dintre vechile noastre probleme ne apar într-o nouă lumină.

Una dintre cele mai tulburătoare şi imediate trăsături ale conştiinţei este *trecerea timpului*. Este ceva atât de familiar, încât e aproape un şoc să aflăm că nici una dintre minunat de precisele noastre teorii privind comportarea lumii reale n-au avut, până acum, nimic de spus despre acest subiect. Chiar mai rău decât atât, ceea ce spun cele mai bune dintre teoriile noastre este în flagrantă contradicţie cu ceea ce ne spun percepţiile noastre despre timp.

Conform teoriei relativităţii generalizate, 'timpul' este doar o alegere particulară de coordonate în descrierea locaţiei unui eveniment spaţio-temporal. Nu există nimic în descrierile spaţio-temporale pe care le dă fizicianul care să singularizeze 'timpul' ca pe ceva ce 'curge'. Într-adevăr, fizicienii consideră adesea modele de spaţiu-timp în care există doar o singură dimensiune spaţială, adăugată unei dimensiuni temporale; în astfel de spaţii doi-dimensionale nu se poate

spune care dintre axe reprezintă timpul şi care spaţiul (vezi figura 7.16). Şi totuşi, nimeni nu va considera că spaţiul 'curge'! Este adevărat că anumite probleme fizice iau în considerare evoluţiile temporale, caz în care suntem interesaţi să calculăm viitorul din starea prezentă a sistemului (vezi paragraful 4.2). Dar aceasta nu este câtuşi de puţin o procedură necesară, iar calculele sunt efectuate în mod normal în acest fel deoarece suntem în general preocupaţi de modelarea matematică a experienţei pe care o avem în lume în termenii 'curgerii' timpului şi pentru că suntem interesaţi să prezicem viitorul [10]. Experienţele noastre curente îşi exercită efectul asupra modelelor noastre calculabile ale lumii şi astfel vorbim de evoluţii (şi le calculăm) pe când legile fizicii nu conţin un asemenea punct de vedere tendenţios şi antropocentric.

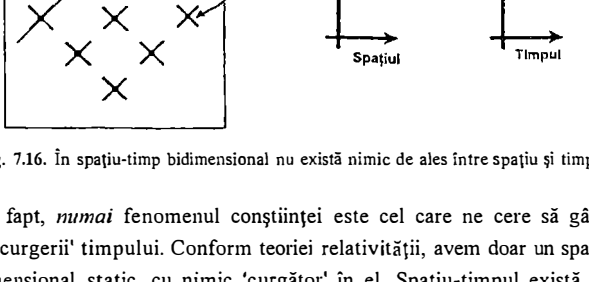


Fig. 7.16. În spaţiu-timp bidimensional nu există nimic de ales între spaţiu şi timp!

De fapt, *numai* fenomenul conştiinţei este cel care ne cere să gândim în termenii 'curgerii' timpului. Conform teoriei relativităţii, avem doar un spaţiu-timp patru-dimensional static, cu nimic 'curgător' în el. Spaţiu-timpul există acolo şi timpul nu 'curge' mai degrabă decât spaţiul. Doar conştiinţa este cea care pare să aibă nevoie ca timpul să curgă, astfel încât nu ar trebui să fim surprinşi să descoperim că relaţia dintre conştiinţă şi timp este ciudată şi în alte feluri.

Într-adevăr, ar fi neînţelept să identificăm prea tare fenomenul conştiinţei conştiente cu aparenta 'curgere' a timpului, ca şi faptul că fizicianul foloseşte parametrul t număr real pentru notaţia a ceea ce se numeşte coordonată temporală. În primul rând, relativitatea ne spune că nu există nici un fel de unicitate privind alegerea paramentrului t, dacă vorbim de spaţiu-timp ca întreg. Sunt posibile mai multe alternative mutual incompatibile, fără nimic special pentru a decide între una şi alta. În al doilea rând, conceptul precis de 'număr real' nu este complet relevant

pentru percepţia noastră conştientă privind trecerea timpului, chiar şi numai pentru motivul că nu avem o percepţie suficient de sensibilă pentru intervale de timp foarte mici să spunem de ordinul unei sutimi de secundă, de exemplu pe când intervalele de timp 'măsurabile' ale fizicienilor pot fi chiar de ordinul a 10^{-25} secunde (aşa cum se vede din acurateţea electrodinamicii cuantice, teoriei cuantice a câmpurilor electromagnetice în interacţiune cu electroni şi alte particula încărcate), sau chiar până la dimensiunile timpului Planck, de 10^{-43} secunde. Mai mult, conceptul de timp ca număr real al matematicienilor ar cere să nu existe o limită inferioară dincolo de care conceptul să-şi piardă semnificaţia (chiar dacă nu mai este fizic relevant).

Este oare posibil să spunem lucruri mai precise despre relaţia dintre experienţa conştientă şi parametrul t pe care fizicienii îl folosesc ca 'timp' în descrierile lor fizice? Poate exista vreun mod experimental de a testa 'când' are loc 'cu adevărat', în relaţie cu acest parametru, o experienţă subiectivă? Dacă spunem că un eveniment al conştiinţei are loc la un anumit moment de timp determinat, înseamnă asta ceva în mod obiectiv? De fapt, anumite experimente de relevanţă specială pentru această problemă au fost într-adevăr efectuate, însă rezultatele lor au fost enigmatice şi cu implicaţii aproape paradoxale. Am dat o descriere a acestor experimente în **MN**, p. 474, însă cred că este nimerit să le reexaminăm aici.

La mijlocul anilor '70, H.H. Kornhuber şi asociaţii săi (vezi Deeke et.al. 1976) au folosit electroencefalograme (EEG) pentru a înregistra semnalele din creierul câtorva voluntari. Ei au încercat să cronometreze orice activitate cerebrală care ar fi putut fi asociată cu un act al *liberului arbitru* (aspectul *activ* al conştiinţei). Subiecţii erau rugaţi să îndoaie un deget, la diferite intervale de timp, însă astfel încât momentele pe care le aleg pentru a face aceasta să fie rezultatul unei *alegeri complet libere*. Experiemntatorii sperau să poată cronometra activitatea cerebrală asociată acestui act de a 'vrea' să îndoaie degetul. Pentru EEG-uri, semnale semnificative se pot obţine numai dacă se mediază pe un număr considerabil de încercări repetate. Ce s-a găsit a fost surprinzător: potenţialul electric înregistrat părea să crească gradual pornind cu o secundă şi jumătate *înainte* de îndoirea degetului. Poate să înseamna asta că un act de voinţă conştientă are nevoie de o secundă sau mai mult pentru a se pune în acţiune? În ce-l priveşte pe subiect, el era convins că decizia de a îndoi degetul este aproape simultană cu actul

(cu siguranţă nu atât de lungă ca o secundă!). (Trebuie să nu uităm că un timp de reacţie 'programat' ca răspuns la un semnal exterior este mult mai scurt decât atât -- cam a cincea parte dintr-o secundă).

Din aceste experimente pare să rezulte că: (i) actul conştient de *liber arbitru* este o pură iluzie; el este preprogramat în activitatea inconştientă precedentă a creierului; *sau* (ii) există un rol de 'ultim minut' al voinţei, care poate uneori (dar nu întotdeauna) să schimbe decizia luată inconştient cu o secundă mai inainte; *sau* (iii) subiectul vrea cu adevărat să îndoaie degetul cu mai mult de o secundă înainte ca îndoirea degetului să aibă loc, însă percepe în mod greşit, într-un mod consistent, faptul că actul conştient apare la un moment mult ulterior, chiar înainte ca degetul să fie, de fapt, îndoit.

În nişte experimente ceva mai recente, Banjamin Libet şi colegii săi au repetat aceste încercări încercând să cronometreze ceva mai direct actul de îndoire voită a degetului, cerând subiectului să ţină minte indicaţiile ceasului din momentul exact când a luat decizia (Libet, 1990, 1992). Concluziile par să confirme rezultatele precedente, dar şi să se pronunţe împotriva concluziei iii). În ce-l priveşte, Libet pare să favorizeze concluzia ii).

În altă clasă de experimente, Libet şi Feinstein au investigat, în 1979, cronometrarea aspectelor *senzoriale* (pasive) ale conştiinţei. Subiecţii testaţi au acceptat să li se fixeze pe creier electrozi, plasaţi în acele porţiuni care primesc semnale senzoriale din diferite puncte ale pielii. În plus faţă de această stimulare directă, în anumite momente se stimula şi punctul corespunzător de pe piele. Concluzia generală a acestor experiemnte este că e nevoie de ceva de ordinul unei jumătăţi de secundă de activitate neuronală (cu anumite variaţii, depinzând de circumstanţe) înainte ca subiecţii să perceapă în mod conştient vreo senzaţie.

Fiecare dintre aceste experimente nu ar fi paradoxal prin el însuşi, deşi prezintă unele aspecte îngrijorătoare. Poate că deciziile aparent conştiente sunt în realitate luate inconştient ceva mai devreme, cel puţin cu o secundă înainte. Poate că senzaţiile noastre au nevoie de ceva de ordinul unei jumătăţi de scundă de activitate cerebrală pentru a se produce. Însă dacă luăm aceste două descoperiri împreună, ele par să ne ducă la concluzia după care orice acţiune în care un stimul extern duce la un răspuns controlat conştient, apare un timp de întârziere de o secundă, o secundă şi jumătate înainte ca răspunsul să poată apărea. Pentru apariţia conştienţei ar fi deci nevoie de aproxiamtiv o jumătate de secundă iar aceasta

odată în funcţiune, maşinăria înceată a *liberului arbitru* pare să mai aibă nevoie de încă o secundă pentru a se porni.

Sunt răspunsurile noastre conştiente chiar atât de încete? Nu acesta pare să fie cazul în conversaţiile obişnuite, bunăoară.

Dacă acceptăm ii), vom fi obligaţi să concluzionăm că cele mai multe acte de răspuns sunt în întregime inconştiente şi că doar din când în când am putea fi în stare să înlocuim un astfel de răspuns cu unul conştient (dar ne-ar trebui aproxiamtiv o secundă pentru a face acest lucru).

Însă, dacă răspunsul este inconştient, cel conştient nu îl poate înlocui decât dacă ambele au acelaşi timp de instalare (altfel, pe când răspunsul conştient ar fi actualizat, cel inconştient ar fi determinat deja un anumit comportament şi n-ar mai putea fi schimbat!). Astfel, în afara cazului în care răspunsurile conştiente sunt rapide, răspunsul inconştient trebuie şi el să fie unul întârziat cu aproximativ o secundă. Pe de altă perte, să ne amintim că o reacţie inconştientă 'programată' poate apărea mult mai repede în aproximativ a cincea parte dintr-o secundă.

Desigur, am putea avea un răspuns inconştient rapid (o cincime de secundă, să spunem), împreună cu posibilitatea (i), ceea ce s-ar reduce la a spune că sistemul de răspuns inconştient ignoră cu desăvârşire orice actuivitate conştientă care ar putea apărea ceva mai târziu. În acest caz (iar situaţia descrisă de iii) este chiar mai rea) singurul rol pentru conştiinţă într-o conversaţie animată ar fi cel de spectator, şi încă la un play-back al întregii drame.

Aici nu este vorba de o contradicţie. Este posibil ca selecţia naturală să fi produs conştiinţa în mod deliberat doar pentru rolul ei în gândirea deliberată, pe când în orice activitate ceva mai rapidă, ea este cel mult un martor pasiv. Întreaga discuţie din partea întâia se referea, în cele din urmă la tipul de contemplaţie conştientă (înţelegerea matematică) despre care ştie oricine cât este de înceată. Poate că facultatea 'conştiinţă' s-a dezvoltat tocmai pentru o astfel de activitate mentală înceată şi contemplativă, pe când răspunsurile rapide sunt acţiuni complet inconştiente, acompaniate poate de o percepţie conştientă întârziată care nu joacă nici un rol activ.

Este cu siguranţă adevărat că atunci când este vorba de un timp suficient de lung, conştiinţa este cea care intervine în acţiune. Pe de altă parte, trebuie să mărturisesc că nu pot crede în posibilitatea ca ea să nu joace nici un rol în activităţi rapide de tipul conversaţiei obişnuite, tenisului de masă sau al curselor de

motociclete, de exemplu.

Mi se pare că există în toată această discuţie o scăpare esenţială şi aceasta este legată de presupoziţia că o cronometrare precisă a momentului în care se produce experienţa conştientă are sens. Să existe oare un *timp actual* la care experienţa conştientă are loc, timp care să preceadă momentul apariţiei oricărui răspuns al liberului arbitru privind experienţa conştientă în discuţie? Luând în considerare relaţia anormală pe care conştiinţa o are cu noţiunea fizică de timp (şi pe care am descris-o la începutul acestui capitol) îmi pare cel puţin plauzibil faptul că nu există un astfel de 'timp' determinat la care trebuie să apară un eveniment conştient [11].

Cea mai cuminte posibilitate în aceste condiţii ar fi o non-localizare difuză în timp, astfel încât să existe o complexitate inerentă în relaţia dintre experienţa conştientă şi timpul fizic. Părerea mea ar fi însă că este vorba de ceva mult mai subtil şi misterios.

Dacă conştiinţa este un fenomen ce nu poate fi înţeles în termeni fizici fără un *input* esenţial din partea teoriei cuantice, s-ar putea întâmpla să existe o interferenţă între misterele **Z** şi concluziile noastre chestionabile privind cauzalitatea, nonlocalizarea şi proprietăţile contrafactuale care pot exista la graniţa dintre conştienţă şi liberul arbitru. De exemplu, poate că tipul de contrafactualitate care joacă un rol în cazul problemei de testare a bombelor din paragrafele 5.2 şi 5.9 ar putea fi important: simplul fapt că un anumit act sau gând *poate* urma să aibă loc, chiar dacă nu are loc în mod actual, poate influenţa comportamentul (ceea ce poate invalida deducţiile logice de tipul celor care exclud posibilitatea ii) de mai înainte).

La modul general, trebuie să avem mare grijă când tragem concluzii logice privind ordinea temporală a evenimentelor, atunci când sunt implicate efecte cuantice (consideraţiile EPR din secţiunea următoare o vor demonstra, dacă mai era nevoie).

Reciproca este că *dacă*, în anumite manifestări ale conştiinţei, raţionamentul clasic, privind ordinea temporală a evenimentelor, ne duce la o concluzie contradictorie, atunci acest lucru reprezintă o puternică indicaţie în favoarea faptului că avem de-a face cu acţiuni cuantice!

7.12. EPR şi timpul: necesitatea unei noi imagini a lumii

Există motive să fim suspicioşi la adresa noţiunii noastre fizice despre timp nu numai în relaţie cu conştiinţa, ci chiar în relaţie cu fizica însăşi, cel puţin atunci când nonlocalizarea cuantică şi contrafactualitatea sunt implicate. Dacă luăm în considerare o interpretare puternic 'realistă' a vectorului de stare $|\psi>$ în situaţiile EPR şi am argumentat pe larg în paragrafele 6.3 şi 6.5 de ce ajungem să avem mari dificultăţi dacă *nu* procedăm astfel — atunci ne aflăm în faţa unei profunde enigme. Astfel de enigme duc la dificultăţi reale privind orice teorie detaliată a naturii teoriei GRW descrisă în paragraful 6.9, sau a tipului de schemă OR din 6.12.

Amintiţi-vă dodecaedrele magice din paragraful 5.3 şi explicaţia lor din paragraful 5.18. Întrebarea este: care dintre următoarele două posibilităţi reprezintă 'realitatea' lucrurilor? Acţiunea colegului meu de a apăsa pe buton este cea care reduce instantaneu (delocalizând) starea originală legată astfel încât starea atomului din dodecaedrul meu este *creată instantaneu*, nelegată, de către apăsarea *lui* pe buton, şi *această* stare redusă defineşte posibilităţile care pot rezulta din acţiunea mea de a apăsa pe bunton? Sau apăsarea *mea* pe buton este cea care primează, acţionând asupra stării legate, pentru a reduce instantaneu starea atomului din dodecaedrul colegului meu, astfel încât el este cel care întâlneşte starea nelegată? Câtă vreme rezultatele sunt cele care ne interesează, nu contează în ce fel privim problema, aşa cum am observat şi în paragraful 6.5. Şi e foarte bine că nu contează, pentru că altfel ar viola principiile relativităţii privind noţiunea de 'simultaneitate', care nu poate avea efecte fizice observabile pentru evenimente aflate la distanţă (separate spaţial). Totuşi, dacă credem că $|\psi>$ reprezintă *realitatea*, aceasta trebuie să fie diferită pentru cele două cazuri anterioare. Unii consideră acest lucru ca pe un motiv suficient pentru a renunţa la interpretarea realistă a lui $|\psi>$. Alţii vor accepta celelalte motive puternice în favoarea interpretării realiste (vezi paragraful 6.3) — şi vor fi dispuşi să amendeze imaginea einsteniană asupra lumii.

Preferinţele mele merg în direcţia salvării ambelor: realism cuantic *şi* esenţa imaginii relativiste a spaţiu-timpului. Dar pentru a face asta, avem nevoie de o schimbare radicală a modului în care reprezentăm realitatea fizică. În loc să insistăm

ca felul în care descriem starea cuantică (sau chiar spaţiu-timpul însuşi) să urmeze procedurile familiare, ar trebui să căutăm un echivalent matematic, dar care ar putea, cel puţin la prima vedere, să arate foarte diferit.

De fapt, există un precedent pentru toate acestea. Înainte ca Einstein să descopere relativitatea generalizată, toată lumea era în extaz în faţa teoriei newtoniaene a gravitaţiei, minunat de exactă, în care particulele, mişcându-se într-un spaţiu plat, se atrăgeau una pe alta conform legii atracţiei gravitaţionale. Am fi putut crede că orice schimbare introdusă în această descriere va distruge remarcabila precizie a schemei newtoniene. Ori, tocmai o astfel de schimbare fundamentală a adus Einstein. În tratarea alternativă a dinamicii gravitaţionale, imaginea a fost complet transformată. Spaţiul nu mai este plat (şi nici măcar nu mai este spaţiu, este spaţiu-timp); nu mai există forţă gravitaţională, ci a fost înlocuită cu efectul de deformare al curburii gravitaţionale. Iar particulele nici măcar nu se mai mişcă, ci sunt reprezentate prin curbe 'statice' în spaţiu-timp.

A fost precizia remarcabilă a teoriei newtoniene distrusă? Chiar deloc; a fost chiar îmbunătăţită şi încă într-un fel extraordinar! (vezi paragraful 4.5)

Am putea oare aştepta ceva similar pentru teoria cuantică? Cred că este foarte probabil. Este nevoie de o schimbare *profundă*, de aceea nu este uşor să vorbim acum despre natura ei specifică. În plus, va părea cu siguranţă prea îndrăzneaţă.

Pentru a încheia acest capitol, voi menţiona două idei care par destul de nebuneşti, deşi nici una destul pentru a fi un candidat suficient în schimbarea discutată. Prima idee se datorează lui Yakir Aharonov şi Lev Vaidman (1990), Costa de Beauregard (1989) şi Paul Werbos (1989). După aceşti autori, realitatea cuantică ar putea fi descrisă prin *doi* vectori de stare, unul propagându-se înainte în timp, de la ultima apariţie a reducţiei **R**, în vreme ce celălalt vector de stare se propagă *înapoi* în timp, de la următoarea ocurenţă a reducţiei **R** aflată în viitor.

Cel de al doilea vector de stare[9] se comportă 'teleologic' în sensul că este guvernat de ceea ce urmează să se întâmple în viitor, în loc de ceea ce s-a întâmplat deja în trecut – o trăsătură pe care mulţi vor găsi-o inacceptabilă. Însă implicaţiile

---

[9] Există un substrat matematic în asocierea vectorului de stare evoluând înapoi un 'bra-vector' $\langle\phi|$, în timp ce vectorului ce evoluează normal (înainte) i se asociază 'vectorul-ket' $|\psi\rangle$. Perechea de vectori de stare poate fi reprezentată ca un produs $|\psi\rangle\langle\phi|$. Aceasta conduce la notaţia cu matricea-densitate din 6.4.

teoriei sunt exact aceleaşi cu ale teoriei standard, şi deci nu poate fi eliminată doar pe considerente de această natură. *Avantajul* ei faţă de teoria standard este că ne permite o descriere obiectivă completă a stării din situaţiile EPR, care pot fi reprezentate în funcţie de spaţiu timp consistent în spiritul relativităţii einsteiniene. Astfel, ne furnizează un fel de soluţie la enigma prezentată la începutul acestei secţiuni însă cu preţul unei comportări teleologice a stării cuantice, pe care mulţi o găsesc îngrijorătoare.

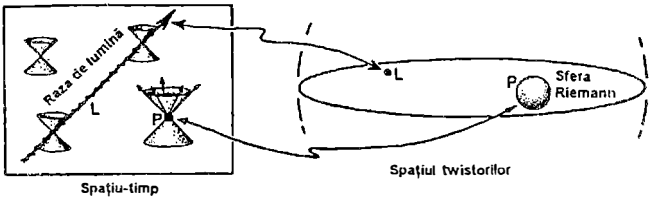


Fig. 7.17. Teoria *twistor*-ilor oferă o interpretare alternativă faţă de spaţiu-timp, în care fiecare rază luminoasă este reprezentată ca un punct, iar evenimentele - prin sfere Riemann.

A doua idee pe care vreau s-o menţionez este *teoria twistorilor* (vezi figura 7.17). Această teorie este motivată tot de enigmele EPR, dar nu oferă (până acum) o soluţie pentru acestea. Tăria ei stă în altă parte: în faptul că oferă descrieri matematice neaşteptate şi elegante unor noţiuni fizice fundamentale (ca, de exemplu, ecuaţiile electromagnetice ale lui Maxwell, date în paragraful 4.4, pentru care este capabilă să dea o formulare matematică atractivă). Furnizează o descriere nonlocală a spaţiu-timpului, în care raze de lumină întregi sunt reprezentate ca simple puncte. Această nonlocalitate spaţio-temporală este cea care se leagă de nonlocalitatea cuantică din paradoxurile EPR. Este de asemenea fundamentată pe *numere complexe* şi geometria asociată acestora, astfel încât obţinem o relaţie strânsă între teoria cuantică U, cu numere complexe şi structura spaţio-temporală. În particular, sfera Riemann din paragraful 5.10 joacă un rol fundamental în relaţia cu conul de lumină al unui punct din spaţiu-timp (şi de pe 'sfera celestă' pentru un observator al acestui punct); vezi David Peat (1998), pentru o descriere a ideilor relevante, sau cartea lui Stephen Huggett şi Paul Tod (1985) pentru o scurtă descriere tehnică [12].

Ar fi nepotrivit să intru în detalii aici. Am menţionat aceste două sugestii

doar pentru a arăta că există mai multe posibile direcţii dacă dorim să schimbăm imaginea noastră despre lumea fizică în ceva diferit de descrierile pe care le susţinem astăzi. Ca o importantă cerinţă de consistenţă pentru o astfel de schimbare, ar trebui să fim în stare să folosim noua descriere pentru a reproduce toate rezultatele de succes ale teoriei cuantice **U** (şi ale relativităţii generalizate). Dar ar fi de asemenea important să mergem mai departe decât atât şi să incorporăm modificările fizice necesare ale teoriei cuantice, în care procedura **R** este înlocuită de un proces fizic real. Cel puţin, aceasta este convingerea mea adâncă; în plus, mai cred, deocamdată, că această 'modificare necesară' trebuie să fie ceva de tipul ideilor descrise în paragraful 6.12. Trebuie menţionat că teoriile în care relativitatea este combinată cu o reducţie a vectorului de stare 'realistă' ca, de exemplu, GRW, s-au lovit de probleme încă insolubile, în particular legate de conservarea energiei. Acest lucru tinde să-mi confirme părerea conform căreia avem nevoie de o schimbare fundamentală în felul în care privim lumea, înainte de a putea fi în stare să facem vreun progres în aceste chestiuni punctuale din fizică.

Cred, de asemenea, că orice progres real în înţelegerea fizică a fenomenului *conştiinţei* va avea de asemenea nevoie – ca o condiţie preliminară – de aceeaşi schimbare fundamentală a imaginii noastre despre lume.

Note şi referinţe bibliografice

1. Vezi, spre exemplu, Lisboa (1992)
2. French (1940), Gelber (1958), Applewhite (1979), Fukui and Asai (1976)
3. Dryl (1974)
4. Hameroff and Watt (1982), Hameroff (1987), Hameroff et al. (1988)
5. Vezi Koruga et al. (1993), pentru introducere a *clatrinilor*, iar Curl şi Smalley (1991), pentru o prezentare populară a fulerenelor
6. Vezi Stretton et al. (1987)
7. Spre exemplu, durata dedusă de Hameroff pentru comutare la dimeri tubulinici pare să corespundă cu frecvenţa lui Fröhlich de circa 5×10^{10} Hz
8. Vezi, spre exemplu, Isham (1989, 1994), Smolin (1993, 1994)
9. Această idee apare într-un manuscris mai vechi al lucrării lui Deutsch (1991), dar nu apare în lucrarea tipărită. David Deutsch m-a asigurat că motivul pentru care a scos

acel pasaj din versiunea finală nu s-a datorat decât opiniei că respectiva idee nu era relevantă pentru scopul central al lucrării respective. În orice caz, după părerea mea, valoarea ideii nu rezidă în corectitudinea pentru orice context existent din gravitaţia cuantică — căci de fapt nu există până acum un astfel de cadru consistent — dar ea poate fi sugestivă pentru dezvoltările ulterioare, şi desigur este! Pentru o interpretare alternativă a non-computabilităţii în *calculul cuantic* vezi Castagnoli et al. (1992)

10. In orice caz, reprezentările noastre fizice normale ale timpului nu disting 'scurgerea' spre viitor, de 'scurgerea' spre trecut (desigur, din cauza celei de a doua legi a termodinamicii, 'reluarea trecutului' nu pare ceva care să fie efectiv realizat prin evoluţia timpului din ecuaţiile dinamicii).
11. Vezi şi Dennett (1991). Unii, care au văzut filmul *A Brief History of Time*, despre Stephen Hawking şi opera sa, poate că au ajuns la un punct de vedere ciudat în privinţa opiniei mele relativ la relaţia conştiinţei cu curgerea timpului. Aş vrea să subliniez cu această ocazie că asta s-a datorat unor nefericite lipsuri şi tăieturi din filmul menţionat.
12. Pentru informaţii mai ample despre *twistori*, vezi şi Penrose şi Rindler (1986), Ward şi Wells (1990), Bailey şi Baston (1990).

8. IMPLICAȚII ?

8.1. 'Aparate' artificiale inteligente

Ce putem deci concluziona, din discuția de mai sus, privind potențialul inteligenței artificiale? Argumentele din partea întâi susțin că tehnologia unor roboți controlați de calculator nu este în stare să furnizeze o cale înspre construcția artificială a unei mașini inteligente — adică al unei mașini care înțelege ceea ce face și poate acționa asupra acestor conținuturi înțelese. Calculatoarele electronice au fără îndoială importanță în clarificarea multor probleme legate de fenomenele mentale (poate, într-o mare măsură, pur și simplu pentru că ne învață ce *nu sunt* fenomenele mentale), și asta fără să mai socotim rolul lor de auxiliar în mai toate problemele de care se leagă progresul științific, tehnologic și social. În concluzie, calculatoarele fac ceva foarte diferit de ceea ce facem noi atunci cînd atenția noastră se îndreaptă asupra unei probleme.

Totuși, ar trebui să fie clar din discuțiile din partea a doua că nu argumentez împotriva posibilității de construcție a unei mașinării inteligente, atâta timp cât prin asta nu înțelegem un dispozitiv controlat computațional. În loc, ar trebui să ne imaginăm mecanisme care să funcționeze prin aceleași tipuri de acțiune fizică ca cea responsabilă de producerea conștienței umane. Cum nu avem nici o teorie fizică pentru această acțiune, este desigur prematur să speculăm dacă sau când un astfel de aparat ar putea fi construit. Cu toate acestea, posibilitatea construcției este deschisă, dacă ne situăm în interiorul punctului de vedere $\mathcal{C}$ în

favoarea căruia am argumentat (conform paragrafului 1.3) și care ne permite să înțelegem mintea în termeni științifici, chiar dacă non-computaționali.

Nu văd nici un motiv pentru care un astfel de aparat să fie biologic. Nu cred că există o linie clară de demarcație între biologie și fizică (sau între biologie, chimie și fizică). Sistemele biologice tind într-adevăr să aibă o organizare mult mai complexă și mai subtilă decât cel mai sofisticat dintre sistemele noastre fizice. Dar, pe de altă parte, înțelegerea noastră fizică este încă la începuturi — mai ales în relație cu fenomenele mentale. Astfel, trebuie să ne așteptăm la creșterea substanțială a sofisticării construcțiilor noastre fizice în viitor. Putem anticipa faptul că această sofisticare viitoare ar putea include efecte fizice care nu pot fi deocamdată percepute.

Nu văd nici un motiv să ne îndoim că, în viitorul imediat, unele dintre misterele mecanicii cuantice (misterele Z) își vor găsi aplicații surprinzătoare. Deja există idei privind aplicarea efectelor cuantice în criptografie pentru a obține rezultate de care aparatele clasice nu sunt în stare. În particular, există anumite propuneri teoretice privind felul în care se poate trimite o informație secretă de la o persoană la alta în așa fel încât dacă este interceptată de o a treia parte, acest lucru să poată fi detectat (C.Benett *et.al.*, 1983). Unele aparate experimentale au fost deja construite și s-ar putea ca ele să-și găsească aplicație comercială în câțiva ani. Numeroase alte scheme care folosesc efectele cuantice au fost propuse în domeniul general al criptografiei, astfel încât domeniul embrionic al *criptografiei cuantice* se dezvoltă rapid. Mai mult, într-o zi, ar putea deveni posibilă construirea unui *calculator cuantic*. Deocamdată, astfel de construcții teoretice sunt încă departe de realizarea practică și este greu de prognozat când ar putea fi ele construite cu adevărat (vezi Obermayer, 1988, a,b).

Este chiar mai greu de prezis posibilitatea construirii unui aparat bazat pe o teorie fizică pe care încă nu o știm. Eu pretind că avem nevoie de o asemenea teorie pentru a putea înțelege fizica aflată la baza unor mecanisme de acțiune non-computațională (în sensul în care am explicat termenul până acum). Conform argumentelor de până acum, pentru a construi un astfel de mecanism avem nevoie în primul rând să găsim o teorie fizică adecvată pentru reducția vectorului de stare — și este greu de spus cât de departe suntem de o astfel de teorie. Este de asemenea posibil ca natura specifică a teoriei OR să ne ofere ea însăși neașteptata complexitate de care avem nevoie.

Până acum *am presupus* că avem nevoie să găsim mai întâi teoria pentru a putea construi o astfel de maşină non-computaţională. Dar s-ar putea să nu fie aşa: preactica experiemntală ne dă numeroase exemple de efecte fizice surprinzătoare care au fost descoperite cu mult înainte de a li se putea da o explicaţie. Un exemplu bun este superconductivitatea, observată experimental cu cel puţin 50 de ani înainte ca Bardeen, Cooper şi Schriffer (1957) să ofere o teorie explicativă completă bazată pe mecanica cuantică. Mai mult, superconductivitatea la temperaturi înalte a fost descoperită în 1986, conform cu Sheng *et.al* (1988), din nou fără motive temeinice de a crede în posibilitatea ei teoretică (şi nici măcar astăzi, în 1994, nu există o explicaţie teoretică adecvată a fenomenului). Pe de altă parte, în cazul activităţii non-computaţionale, ar fi greu de văzut cum ar putea cineva *spune* că un obiect oarecare se comportă non-computaţional. Conceptul de computabilitate este mai aproape de teorie decât observaţional. Dar, în interiorul unei teorii non-computaţionale, ar putea foarte bine să existe o comportare testabilă, pe care un mecanism actual ar putea-o avea. În concluzie, opinia mea este că e foarte puţin probabil să putem observa un comportament non-computaţional într-un obiect fizic construit în absenţa unei teorii preliminare.

Pentru a putea continua discuţia, să încercăm să ne imaginăm, însă, că *avem* teoria fizică necesară — care, aşa cum am argumentat, trebuie să fie o teorie non-computaţională OR a reducţiei vectorului de stare; şi să presupunem de asemnea că avem o confirmare experimentală a acestei teorii. Cum putem proceda pentru a contrui un mecanism *inteligent*? Răspunsul este că *nu putem* — cel puţin nu doar pe baza acestei teorii. Avem nevoie de o altă relizare teoretică, o realizare care să ne spună cum apare conştiinţa ca rezultat al unei organizări corespunzătoare, în care efectele OR non-computaţionale îşi au locul lor. Eu, cel puţin, n-am nici un fel de idee cum ar putea arăta o astfel de dezvoltare teoretică. După modelul exemplelor de mai sus cu superconductivitatea, am putea să ne imaginăm că un aparat de tipul căutat ar putea fi descoperit accidental, fără a avea mai întâi o teorie a conştiinţei. Nu mai e nevoie să spunem că acest lucru este extrem de improbabil — în afară de cazul în care luăm în considerare un proces de evoluţie conform căruia inteligenăa poate apărea pur şi simplu datorită beneficiilor directe pe care conştiinţa le conferă, fără ca din partea noastră să existe vreo înţelegere privind felul în care a fost obţinut acest lucru (ceea ce este exact ce se întâmplă cu noi!). Tot acest proces va fi cu siguranţă foarte lent — în special

dacă luăm în considerare cât de mult timp îi trebuie conștiinței pentru a-și manifesta avantajele! Cititorul ar putea ajunge la concluzia că un mod mult mai satisfăcător de a construi aparate inteligente este de a adopta procedurile aleatoare dar remarcabil de efective care sunt deja la lucru de milenii întregi!

Desigur că nici unul dintre aceste lucruri nu ne va opri din încercarea de a vedea ce anume se petrece în conștiința și în inteligența noastră. Și eu vreau să cunosc aceleași lucruri. Până la urmă, argumentele acestei cărți arată că ceea ce *nu* se petrece este mai ales o mare cantitate de activitate computațională, iar ceea ce *se petrece* nu are șansa să fie cu adevărat înțeles până când nu vom avea o cunoaștere mult mai profundă a naturii materiei, timpului, spațiului și a legilor care le guvernează. Avem nevoie de asemenea de o cunoaștere mult mai detaliată a fiziologiei creierului, în particular la nivelele microscopice care s-au bucurat de o atenție deosebită în anii din urmă. Va trebui să știm mai multe privind circumstanțele în care conștiința apare sau dispare, privind problema curioasă a timpului în care ea se manifestă, a modului în care este folosită și a avantajelor specifice pe care posedarea unui astfel de obiect le poate aduce – și asta în plus față de alte multe probleme în care testarea obiectivă este posibilă. Este un câmp de cercetare extrem de larg și deci putem anticipa în el progrese în mai multe direcții diferite.

8.2. Lucruri pe care calculatoarele le fac bine – sau prost

Chiar dacă acceptăm că tipul teoretic de calculator actual *nu* va ajunge niciodată să obțină inteligență sau conștiență, ne aflăm în continuare în fața puterii și a performanțelor extraordinare pe care calculatoarele moderne le pot dezvolta, cu perspectiva unei creșteri masive în viitor (vezi paragrafele 1.2, 1.10 și Moravec, 1988). Cu toate că aceste mașini nu vor *înțelege* lucrurile pe care le fac, vor continua să le facă, incredibil de repede și de precis. Va fi o astfel de activitate chiar fără mintea asociată în stare să obțină lucruri pentru obținerea cărora noi ne folosim mințile, poate chiar mai eficient decât putem s-o facem noi? Putem ajunge să ne imaginăm tipurile de lucruri pe care un calculator va ajunge să le facă foarte bine, sau cele pe care mintea umană le va face întotdeauna mai bine?

Deja calculatoarele pot juca şah extraordinar de bine, apropiindu-se de nivelul celor mai mari maeştri umani. Totuşi, la jocul de GO, ele par să nu fi ajuns prea departe. Atunci când jocurile de acest fel trebuie jucate foarte repede, acesta este un atu în favoarea calculatoarelor, pe când timpul îndelungat de joc favorizează jucătorii umani. Problemele de şah care presupun anticiparea a una sau două mutări pot fi rezolvate aproape instantaneu de calculator, indiferent cât de grele le găseşte un om. Pe de altă parte, o problemă cu o idee simplă, dar care cere, să zicem, 50 sau 100 de mutări, poate înfrânge în mod radical calculatorul, pe când jucătorul cu experienţă o rezolvă fără mare dificultate (vezi paragraful 1.15, Fig.1.7).

Aceste diferenţe pot fi înţelese, în mare, în funcţie de anumite distincţii între problemele la care calculatoarele şi oamenii sunt buni. Calculatorul efectuează doar calcule, fără să înţeleagă ce face — cu toate că foloseşte o parte din înţelegerea *programatorilor* săi. El poate înmagazina o mare cantitate de informaţie, dar şi jucătorul uman poate face asta. Calculatorul poate aplica foarte rapid şi complet mecanic o mare cantitate de secvenţe — aplicaţii ale înţelegerii programatorului — însă face acest lucru cu o abilitate care depăşeşte cu mult posibilităţile unei fiinţe umane. Jucătorul uman trebuie să rejudece şi să reinventeze strategii în funcţie de situaţie, în plus faţă de înţelegerea generală privind regulile jocului. Acestea sunt calităţi pe care calculatorul nu le poate avea, însă el îşi poate folosi în mare măsură puterea sa de calcul pentru a compensa aceste deficienţe.

Să presupunem că numărul de posibilităţi pe mutare pe care calculatorul trebuie să le ia în considerare este, în medie, p; atunci, pentru o secvenţă de m mutări, trebuie luate în considerare p^m alternative. Dacă calcularea fiecărei alternative ia un timp mediu t, atunci avem ceva de genul:

$$T = t \times p^m$$

pentru timpul total T necesar pentru a calcula strategia mutării. În cazul jocului de GO, numărul de mutări care trebuie luate în considerare este mult mai mare, de ordinul $p = 200$, astfel încât un calculator nu poate lucra cu m mai mare decât 5 sau ceva de acest gen. Evident, judecăţile jucătorului uman sunt mult mai lente decât ale calculatorului (t mare pentru om, mic pentru calculator) însă, ele pot reduce numărul p la numărul mutărilor efective care trebuie luate în considerare (p efectiv mic pentru om, mare pentru calculator), deoarece doar puţine alternative vor fi demne de a fi luate în consideraţie.

Rezultă, la modul general, că jocurile în care p este mare dar poate fi redus semnificativ prin folosirea înţelegerii şi a judecăţii, sunt în avantajul jucătorului uman. Pentru că, dându-se un T destul de mare, actul omului de a 'reduce efectiv' numărul p, are mult mai mare importanţă în obţinerea unui m mare în formula $T = t \times p^m$ decât dacă facem timpul t foarte mic (operaţie la care calculatorul este bun). Însă pentru T mic, reducerea lui t este cea care are sens. Aceste lucruri sunt consecinţe simple ale formei 'exponenţiale' a expresiei $T = t \times p^m$

Toate aceste consideraţii sunt puţin cam schematice, însă cred că punctul esenţial este destul de clar. (Dacă nu sunteţi matematician şi vreţi să aveţi o imagine despre comportarea lui $t \times p^m$, încercaţi câteva exemple de valori t,p şi m.) Nu merită să intrăm în prea multe detalii aici, însă o clarificare poate fi utilă. Se poate argumenta că o anticipare de lungă durată, măsurată cu m, nu este scopul pentru care se luptă jucătorul uman. Însă, de fapt, exact asta este. Atunci când jucătorul judecă poziţiile la nivelul câtorva mutări înainte, şi apoi consideră că nu este de folos să calculeze mai departe, aceasta este un calcul efectiv pe mult mai multe mutări înainte, deoarece judecata umană cuprinde efectul probabil al mişcărilor care urmează a fi făcute. În orice caz, pe baza unor consideraţii sumare de acest fel este posibil să obţinem o anumită înţelegere a faptului că este mai uşor pentru calculatoare să rezolve probleme de şah scurte decât să *participe* la partide complete, la jocuri lungi, este mai uşor să joace şah decât GO şi au un avantaj (relativ) la jocul rapid.

Aceste argumente nu sunt prea sofisticate, însă principalul este că ingredientul esenţial care-i lipseşte calculatorului este *judecata*, bazată pe *înţelegere* (vezi şi figura 1.7 în paragraful 1.15). Înţelegerea conştientă este un proces comparativ încet, însă care poate reduce considerabil numărul de alternative care merită luate în considerare, şi deci poate creşte considerabil adâncimea calculului. (Alternativele nu trebuie nici măcare încercate, dincolo de un anumit punct.) De fapt, mi se pare că dacă considerăm performanţele pe care calculatoarele le-ar putea atinge în viitor, o bună călăuză pentru un răspuns ar fi obţinută din întrebarea „este înţelegerea necesară pentru efectuarea sarcinii?" Mute lucruri din viaţa de zi cu zi nu au nevoie de înţelegere pentru a fi efectuate, şi este foarte posibil ca roboţii construiţi de calculator să devină foarte buni pentru astfel de lucruri. Deja există maşini controlate de reţele neuronale artificiale care efectuează sarcini extrem de importante de acest tip. De exemplu, ele se pot

descurca la recunoaşterea feţelor, pot executa prospecţiuni mineralogice, pot recunoaşte defectele dintr-o maşinărie prin distingerea între diferite sunete, pot identifica fraude bancare etc [1]. La modul general, aceste metode au succes; maşinile pot atinge şi chiar depăşi capacitatea experţilor umani. Dar programarea *bottom-up*, nu conduce la tipul de 'expertiză' care apare în cazul sistemelor *top-down*, să spunem cu calculatoarele care joacă şah sau, chiar mai impresionant, în cazul calculului numeric, unde cei mai buni calculatori umani nu pot ajunge nici pe departe la performanţele unui calculator electronic. În cazul sarcinilor în care se lucrează cu reţele neuronale artificiale, trebuie să spunem că nu ştim prea clar cum rezolvă oamenii aceste sarcini. În cazurile în care este vorba de o bună organizare *top-down* în programarea calculatorului, ca de exemplu în cazul calculului numeric, al calculatoarelor care joacă şah etc., eficienţa calculatorului este foarte mare. În aceste cazuri, calculatorul nu are nevoie de nici un fel de înţelegere actuală a sarcinii pe care o are de efectuat, deoarece înţelegerea relevantă a fost furnizată de către programatorii umani (vezi paragraful 1.21).

Trebuie de asemenea menţionat că, foarte frecvent, în sistemele *top-down* apar erori de calcul, pentru că programatorul face greşeli. Evident, este rezultatul unei erori umane şi nu are nimic de-a face cu neajunsurile calculatorului.

Tipul de situaţie în care poate fi periculos să ne punem toată încrederea într-un sistem complet controlat de un calculator este atunci când sistemul poate executa o sarcină corect vreme îndelungată, poate chiar dând impresia că *înţelege* ce face. Apoi, complet pe neaşteptate poate face ceva complet nebunesc, demonstrând că nu a avut nici un moment vreo înţelegere reală a situaţiei (este cazul eşecului calculatorului Deep Thought cu poziţia de şah din figura 1.7). Astfel, trebuie să fim mereu precauţi realizând că 'înţelegerea' nu este o calitate computaţională.

Desigur, oamenii înşişi diferă foarte mult în ce priveşte calitatea înţelegerii. Ca şi în cazul calculatoarelor, se poate întâmpla ca un om să ne dea impresia că a înţeles fără să aibă cea mai mică idee despre ce e vorba. În plus, există un joc permanent între înţelegerea reală pe de-o parte, şi memoria şi calculul, pe de alta. Calculatoarele sunt capabile de cele din urmă, nu şi de cea dintâi. Aşa cum este foarte familiar pentru profesorii de pretutindeni (dar nu şi pentru guverne, din păcate) calitatea *înţelegerii* este de departe cea mai importantă. Înţelegerea şi nu repetarea papagalicească a regulilor sau informaţiilor este ceea ce orice profesor

vrea să încurajeze la elevii săi. Într-adevăr, una din cerințele în alcătuirea întrebărilor de examene (mai ales în cazul matematicii) este ca ele să testeze înțelegerea candidatului, ca distinctă de memoria și puterea de calcul deși nici acestea nu sunt calități lipsite de valoare.

8.3. Estetică ... și altele

În discuția de mai sus m-am concentrat asupra calității 'înțelegerii' ca ingredientul esențial care lipsește din orice sistem pur calculatoriu. Această calitate esențială a fost, de fapt, subliniată în argumentul lui Gödel prezentat în paragraful 2.5. Dar înțelegerea este doar una din calitățile pentru care conștiința conștientă are valoare pentru noi. Mai general, ființele conștiente beneficiază de orice circumstanță în care pot 'simți' direct lucrurile; ori *exact aceasta*, spun eu, este ceea ce un sistem computațional nu va obține niciodată.

Ne putem foarte bine întreba în ce fel este un robot controlat de calculator *dezavantajat* de inabilitatea lui de a simți de faptul că nu poate aprecia, să spunem, frumusețea unui cer înstelat, splendoarea magnifică a Taj Mahalului sau complexitatea magică a unei fugi de Bach -- sau chiar frumusețea clară a teoremei lui Pitagora. Am putea spune doar că este pierderea robotului, faptul că nu poate simți ceea ce suntem noi capabili să simțim atunci când suntem confruntați cu asemenea manifestări de calitate. Și totuși este mai mult decât atât. Am putea pune o întrebare diferită. Acceptând că robotul nu este într-adevăr capabil să simtă nimic, n-ar putea, totuși, un calculator inteligent programat să fie capabil să producă mari opere de artă?

E o întrebare delicată, din câte mi se pare. Cred că răspunsul scurt este pur și simplu 'nu' chiar și numai pentru faptul că calculatorul nu posedă calitățile necesare pentru a distinge între bine și rău sau superbul de acceptabil. Dar ne-am putea întreba: ce este *necesar* pentru ca el să 'simtă', pentru a-și dezvolta propriile 'criterii estetice' și pentru a emite propriile sale judecăți? Ne-am putea imagina că astfel de judecăți ar putea să apară la capătul unei lungi perioade de antrenament (*bottom-up*). Cu toate acestea, ca și în cazul înțelegerii, cred că este mult mai probabil ca aceste criterii să fie parte a intrărilor oferite deliberat calculatorului, distilate în urma unei analize *top-down* detaliate (foarte probabil asistată de

calculator) și executate de oameni. Într-adevăr, scheme de acest fel au fost puse în acțiune de un număr de cercetători ai inteligenței artificiale. De exemplu, Christopher Longuet-Higgins, într-o cercetare efectuată la Universitatea din Sussex, a introdus variate programe de compus muzică, conform cu anumite criterii pe care le-a furnizat în prealabil. Încă din secolul optsprezece, Mozart și contemporanii săi au arătat că este posibilă construirea unui 'zar muzical' care ar putea fi folosit pentru a combina ingrediente plăcute estetic cu altele, aleatoare, pentru a produce compoziții (discutabile, e drept). Mecanisme similare au fost folosite în artele vizuale — cum ar fi sistemul 'AARON' programat de Harold Cohen, care poate desena o sumedenie de linii 'originale' prin invocarea unor elemente aleatoare pentru a fixa ingredienții de intrare conform unor anumite reguli (vezi Margaret Boden(1990), *The Creative Computer*; Michie și Johnson, (1984))

Cred că veți accepta că produse de acest gen nu sunt, cel puțin deocamdată, de natură a fi comparate cu produsele unui artist uman, chiar modest. Cred că n-ar fi nepotrivit dacă am spune că ceea ce lipsește din produsul finit este 'sufletul' Cu alte cuvinte, opera nu *exprimă* nimic pentru că calculatorul nu *simte* nimic.

Desigur, din timp în timp, o astfel de 'creație' generată în mod aleator de calculator ar putea avea, din pură întâmplare, un oarecare merit artistic. (Ceea ce nu este lipsit de legătură cu vechea poveste a posibilității generării piesei *Hamlet* prin tastarea literelor la întâmplare.) Într-adevăr, trebuie să admitem că Natura însăși este capabilă să producă multe opere de artă în mod aleator, ca în cazul formării cristalelor sau al cerului înstelat. Însă, în lipsa abilității de a *simți* acea frumusețe, nu există nici o posibilitate de a distinge între frumos și urât. În acest proces de *selecție* își va arăta sistemul computațional fundamentala lui limitare.

Din nou, am putea presupune că este posibilă alimentarea calculatorului cu astfel de criterii și că ele ar putea funcționa mulțumitor pe principiul generării unui mare număr de exemple similare — cel puțin până când producțiile de acest fel ar deveni plictisitoare și ar fi nevoie de ceva nou. În acest punct, judecățile estetice propriu-zise ar deveni necesare pentru a putea decide care dintre 'noile idei' are merit artistic și care nu.

Astfel, pe lângă calitatea *înțelegerii* există și alte calități ce vor lipsi întotdeauna dintr-un sistem computațional; de exemplu cele estetice. Lor trebuie să le adăugăm și alte tipuri de lucruri, cum ar fi judecățile *morale*. Am văzut în partea întâia că judecata asupra ce este sau nu *adevărat* nu se poate reduce la un

simplu calcul. Acelaşi lucru (poate mai evident) se aplică şi în cazul *frumosului*, sau al *binelui*. Acestea sunt chestiuni care au nevoie de conştienţă şi sunt deci inaccesibile roboţilor controlaţi de calculator.

Dincolo de natura lor non-computaţională am putea întreba dacă *frumuseţea* şi *binele* sunt calităţi *absolute*, în sensul platonic în care 'absolut' se foloseşte pentru adevăr mai ales pentru adevărul matematic. Platon însuşi a argumentat în favoarea unui asemenea punct de vedere. Ar putea fi conştienţa noastră capabilă să intre în contact cu asemenea entităţi absolute? Asta să fie ceea ce-i dă conştiinţei o asemenea putere? Poate că aici am putea avea un indiciu asupra a ce anume 'este' conştiinţa şi în virtutea cărui 'scop' Să fie conştienţa cea care întinde o punte de trecere înspre lumea formelor platonice?

8.4. Anumite pericole inerente în tehnologia calculatoarelor

Cu o tehnologie în expansiune, este evident că ne confruntăm şi cu pericole, nu numai cu avantaje. Astfel, pe lângă avantajele evidente pe care calculatoarele ni le oferă, există multe ameninţări potenţiale pentru societatea noastră, inerente în tehnologia calculatoarelor. Una din principalele probleme pare să fie complicaţia extraordianră generată de interconectare cu care calculatoarele ne confruntă. Nu este doar o problemă a calculatorarelor, ci mai ales a comunicării globale care leagă calculatoarele unele cu altele, în jurul planetei noastre. Un bun exemplu este instabilitatea burselor, unde se pot face averi instantanee; dar este foarte probabil ca multe alte probleme să apară doar din cauza complexităţii şi a interconexiunilor sistemului ca întreg.

Bănuiesc că unii ar putea crede că apariţia unor sisteme interconexate prea complexe pentru a putea fi înţelese de oameni *nu* este o problemă serioasă. Aceştia ar putea avea încredere într-o perspectivă în care, în cele din urmă, calculatoarele însele vor câştiga înţelegerea necesară pentru a se descurca cu acest sistem. Am văzut însă că înţelegerea nu este o calitate de care calculatoarele să fie capabile, astfel încât nu există şansa unei soluţii liniştitoare din această parte.

Mai sunt şi alte probleme adiţionale care rezultă din faptul că avantajele tehnologice sunt foarte rapide, astfel încât un model de calculator 'se învecheşte'

foarte repede, fiind mereu depăşit de altele. Rezultatele cerinţei de înnoire continuă şi rapidă, în special a folosirii unor sisteme incomplet testate din cauza presiunii de pe piaţă, pot conduce la o înrăutăţire a situaţiei în viitor.

8.5. Alegerile enigmatice

Data alegerilor mult-aşteptate se apropie. Numeroase sondaje de opinie au fost efectuate în ultimele săptămâni. Într-un mod consistent, partidul de guvernământ conduce cu trei sau patru procente. Evident, există fluctuaţii şi deviaţii de o parte şi de alta a acestei cifre — mai ales pentu că cifrele sondajelor de opinie se bazează pe eşantioane relativ mici, pe când întreaga populaţie votantă are considerabile variaţii de opinie de la un loc la altul. Într-adevăr, marja de eroare pentru aceste sondaje de opinie poate ea însăşi ajunge până la 3% sau 4%, astfel încât nici unul dintre sondaje nu este cu adevărat credibil. Şi totuşi, totalitatea dovezilor este convingătoare. Sondajele luate împreună au o marjă de eroare mult mai mică, iar acordul dintre ele pare să aibă exact felul de uşoară variaţie la care ne-am aştepta pe baze statistice. Rezultatele mediate au o eroare mai mică de 2%. Unii ar putea spune că există o mică deplasare în favoarea partidului de guvernământ în ajunul zilei alegerilor şi că la vot, o mică parte dintre cei nedecişi ar putea fi convinşi să voteze cu partidul la putere. Chiar şi aşa, o deplasare a cifrelor faţă de sondaje în favoarea partidului de guvernământ nu va rezolva nimic dacă la vot acesta nu va câştiga cu aproximativ 8% mai mult decât rivalii cei mai apropiaţi; numai astfel ar putea partidul de guvernământ să blocheze o coaliţie a adversarilor. Până la urmă, sondajele de opinie sunt doar predicţii, nu-i aşa? Doar *adevăratul* vot va exprima voinţa populară, ori aceasta va fi obţinută abia din cifrele votului real, în ziua alegerilor.

Vine şi ziua alegerilor,... şi trece. Voturile sunt numărate, iar rezultatul este o completă surpriză pentru aproape toată lumea — şi mai ales pentru organizatorii de sondaje care şi-au pus la bătaie o mare parte a energiei şi timpului lor, pentru a nu mai menţiona propriile reputaţii. Partidul de guvernământ a câştigat din nou, cu o majoritate confortabilă, realizând diferenţa de 8% faţă de adversarii cei mai apropiaţi. Un mare număr de votanţi sunt uluiţi — chiar oripilaţi. Alţii, chiar dacă complet surprinşi, sunt încântaţi. Şi totuşi rezultatul este

fals. 'Aranjarea' votului s-a realizat prin metode subtile, care au scăpat tuturor observatorilor. N-au existat urne de vot supraumplute, nici una n-a fost pierdută, substituită sau duplicată. Oamenii care au numărat voturile și-au făcut treaba în mod conștiincios și cinstit și, în general, fără greșeli. Și totuși rezultatul este profund greșit. Cum s-a putut întâmpla? Cine este responsabil?

S-ar putea întâmpla ca întregul guvern să ignore cu desăvârșire ce anume s-a întâmplat. Cu toate că sunt beneficiarii acestei înșelătorii, nu trebuie neapărat să fie responsabili. Există alții, în spatele scenei, care s-au temut că în cazul unei înfrângeri a partidului de guvernământ, propria lor existență ar fi în pericol. Ei sunt o parte a unei organizații în care partidul de guvernământ are mare încredere (cu bune motive!), mai mare decât adversarii lui, o organizație strict secretă pe care partidul de guvernământ a ajutat-o să se camufleze și chiar să se extindă. Cu toate că organizația este legală, o mare parte a activității ei nu este, iar printre metodele ei se numără șantajul politic și alte acte ilegale. Poate că membrii ei au o îngrijorare reală (bine întreținută) legată de consecințele unei guvernări a opoziției, care ar putea distruge țara, sau ar putea-o chiar 'vinde' puterilor străine. Iar printre membrii organizației se numără experți de rang înalt în construirea virușilor electronici.

Amintiți-vă ce poate face un virus unui calculator. Printre cei mai cunoscuți sunt aceia care în anumite zile distrug toate datele dintr-un calculator infectat. Se poate întâmpla ca operatorul să privească îngrozit și neajutorat cum literele de pe ecran se deformează, alunecă și dispar, înlocuite eventual de un mesaj obscen. În orice caz, toate datele din memoria calculatorului sunt pierdute. Mai mult, orice dischetă introdusă în calculator se infectează cu virusul respectiv pe care îl va introduce în orice alt calculator, unde va fi folosită. Programele antivirus pot, în principiu, distruge o astfel de infecție, dacă este localizată, însă doar dacă natura virusului este deja cunoscută. După ce virusul a lovit, nu se mai poate face nimic.

Acești viruși sunt construiți în mod normal de hackeri amatori, de obicei programatori frustrați care vor să-și ia revanșa, făcând rău uneori pentru rațiuni detectabile, alteori aparent fără motiv. Însă membrii organizației noastre nu sunt amatori; ei sunt profesioniști, plătiți foarte bine, cu statut și cunoștințe de experți. Probabil că multe din activitățile lor sunt 'onorabile', în întregime în interesul țării; și totuși ei acționează adesea și în scopuri mai puțin cunoscute și care nu au nimic de-a face cu morala. Virușii produși de ei nu pot fi detectați de nici un

program antivirus standard şi sunt programaţi astfel încât să lovească exact în ziua alegerilor (o dată cunoscută cu mult timp înainte liderului partidului de guvernământ şi apropiaţilor săi). După ce şi-au îndeplinit misiunea — o acţiune mult mai subtilă decât distrugerea datelor — virusii sunt programaţi să se autodistrugă, fără să lase nici o urmă a trecerii lor în afara stricăciunilor provocate.

Pentru ca un astfel de virus să fie eficient în alegeri, este necesar ca un anumit stadiu al numărătorii voturilor să fie făcut în întregime de calculatoare, fără intervenţia umană (un virus poate infecta doar un calculator programabil, nu şi calculatoarele de făcut socoteli 'cu mâna'). Conţinutul urnelor de vot este astfel numărat corect, însă rezultatele acestei numărători primare trebuie adunate. Este, desigur, mult mai eficient şi mai uşor ca aceste date să fie adunate de un calculator. Nu va fi nici o eroare. Indiferent ce calculator folosim, vom obţine acelaşi rezultat. Membrii partidului de guvernământ vor obţine aceleaşi rezultate ca şi oponenţii lor, ca şi oricare dintre observatorii neutri etc. Poate că toţi aceştia folosesc mărci diferite de calculatoare, însă asta nu este o piedică. Experţii organizaţiei ştiu care sunt toate aceste mărci şi au produs câte un virus pentru fiecare. Construcţia lor va diferi uşor, însă rezultatele vor fi identice, iar acordul între calculatoare îi va convinge chiar şi pe cei mai înveteraţi sceptici.

Cu toate că acordul între calculatoare este exact, numerele sunt false. Ele au fost modificate de o formulă precisă, care depinde până la un punct de numărul real de voturi — de unde gradul mare de plauzibilitate al rezultatului — astfel încât să-i dea partidului de guvernământ numărul de voturi de care are nevoie; şi chiar dacă credibilitatea este pusă puţin la încercare, rezultatul final poate fi acceptat. Va părea, pur şi simplu, că în ultimul minut, un număr semnificativ de indecişi s-au răzgândit şi au votat pentru partidul de guvernământ.

În situaţia ipotetică pe care o descriu, indecişii nu s-au răzgândit, iar rezultatul a fost trucat. Deşi inspiraţia pentru această poveste mi-a venit în urma alegerilor din Marea Britanie din 1992, trebuie să mă grăbesc să subliniez că sistemul oficial de numărare a voturilor adoptat în Marea Britanie nu permite acest tip de fraudă. Numărătoarea se face 'cu mâna' la toate nivelurile. Deşi pare o metodă învechită şi depăşită, este important să o menţinem — sau, cel puţin, să reţinem un sistem care să ne permită să evităm chiar cele mai mici suspiciuni de fraudă de acest tip.

De fapt, pe de altă parte, calculatoarele ne pot oferi minunate ocazii de a

elabora sisteme de vot în care opiniile electoratului să fie mult mai corect reprezentate decât sunt acum. Nu este momentul să intrăm în asemenea detalii aici, însă punctul esenţial este că e posibil pentru fiecare votant să furnizeze la alegeri mult mai multă informaţie decât simpla decizie în favoarea unui candidat sau a altuia. Cu un sistem controlat de calculator, această informaţie poate fi analizată instantaneu, astfel încât rezultatul poate fi cunoscut imediat ce votul s-a încheiat. Cu toate acestea, aşa cum am încercat să arăt cu povestea de mai sus, ar trebui să fim extrem de precauţi cu asemenea sisteme şi să avem metode şi tehnici de control care să ne protejeze de astfel de situaţii.

Nu doar în cazul alegerilor trebuie să fim precauţi; sabotarea sistemului contabil al unei companii rivale, de exemplu, este o altă situaţie în care poate fi folosită o tehnică de virusare. Şi există nenumărate alte căi pe care viruşii ar putea fi puşi la lucru. Sper că povestea mea este de natură să sublinieze necesitatea unui control uman permanent al autorităţii aparente şi aparent credibile a calculatoarelor. Problema nu e doar că calculatoarele nu înţeleg nimic, ci, mai ales, că sunt foarte uşor de manipulat de acei câţiva, foarte puţini, care stăpânesc la perfecţie modul în care sunt ele programate.

8.6. Fenomenul fizic al conştiinţei?

Scopul părţii a doua a acestei cărţi a fost căutarea, în interiorul unei explicaţii ştiinţifice, a unui loc fizic pentru experienţa subiectivă. Am argumentat de ce cred că o astfel de întreprindere are nevoie de o extindere a înţelegerii ştiinţifice actuale. Personal, nu mă îndoiesc că locul în care trebuie să ne uităm dacă vrem să schimbăm imaginea noastră asupra fizicii actuale este fenomenul reducţiei stării cuantice. Pentru ca fizica să fie în stare să înglobeze ceva atât de fundamental străin tipului de descriere fizică curentă cum este fenomenul conştiinţei, trebuie să ne aşteptăm la o profundă schimbare — o schimbare care să modifice însăşi elemenele constitutive ale imaginii noastre filozofice despre natura realităţii. Voi avea câteva comentarii de făcut asupra acestui punct în secţiunea care va încheia această carte. Pentru moment, să încercăm să punem o problemă mai simplă: unde ar trebui să ne aşteptăm să găsim conştiinţa în lumea cunoscută?

Trebuie să încep prin a sublinia că argumentele pe care le-am prezentat au

foarte puţin de spus la modul pozitiv. Ele spun mai curând că actualele calculatoare nu sunt conştiente, însă au multe de spus depre momentul în care ne-am putea aştepta ca un obiect să fie conştient. Experienţa pare să ne sugereze, cel puţin până acum, că în structurile biologice este cel mai probabil să găsim fenomenul care ne interesează. La un capăt al scalei avem fiinţele umane, pentru care este cu siguranţă clar unde se petrece fenomenul conştiinţei — un fenomen despre care presupunem că este în mod normal prezent în asociaţie cu creierul uman conştient (treaz şi, probabil, în stare de vis).

Ce putem spune despre celălalt capăt al scalei? Am argumentat că locul cel mai probabil în care am putea observa efectele cuantice colective coerente este în microtubulele citoscheletului, mai degrabă decât în neuroni şi că fără o coerenţă cuantică nu vom fi în stare să găsim un rol noii fizici OR, capabilă să ne ofere ingredientul non-computaţional necesar pentru a vorbi despre fenomenul conştiinţei în termeni ştiinţifici. Am văzut însă că citoscheletele se află peste tot, de la celulele eucariote — celule care constituie plantele şi animalele — la organisme unicelulare, ca amoeba sau parameciul, dar nu în bacterii. Ar trebui deci să ne aşteptăm să găsim urme ale unei conştiinţe la parameci? Putem spune că un parameci 'ştie' ce face, oricare ar fi sensul în care am folosi acest cuvânt? Ce putem spune despre celulele umane individuale, din creier, sau poate din ficat? N-am nici cea ma vagă idee dacă vom fi sau nu obligaţi să acceptăm astfel de absurdităţi aparente atunci când înţelegerea naturii fizice a conştienţei ne va permite să răspundem acestor întrebări. Dar, un lucru este ceea ce *cred* eu despre aceste probleme şi, cu totul alta este *întrebarea ştiinţifică* la care va trebui în cele din urmă să răspundem, indiferent cât de departe suntem în prezent de a răspunde la ea.

Se pretinde uneori, pe temeiuri filozofice generale, că nu există nici o modalitate de a şti dacă vreo altă fiinţă în afara mea este sau nu conştientă — şi asta fără să mai introducem parameciul în discuţie. În ce mă priveşte, cred că această direcţie de gândire este exagerat de pesimistă şi prea îngustă. Până la urmă, nu ne interesează niciodată probelemele *certitudinii absolute* atunci când stabilim prezenţa unei calităţi fizice într-un obiect. Nu văd nici un motiv pentru care n-am fi în stare să ajungem la un stadiu în care am putea răspunde la întrebările privind posesia conştiinţei conştiente cu precizia cu care astronomii vorbesc despre corpuri cereşti aflate la mulţi ani-lumină. Nu este atât de mult timp de când compoziţia materială a soarelui şi a stelelor era considerată incognoscibilă, ca şi trăsăturile

feței întunecate a lunii. Cu toate acestea, întreaga suprafață a lunii este acum cartografiată (cu ajutorul navetelor spațiale), iar compoziția soarelui este în mare măsură înțeleasă (prin observarea liniilor spectrale și modelarea detaliată a interiorului coroanei solare). Compoziția detaliată a multor stele este deja cunoscută cu un bun grad de aproximație. Și chiar alcătuirea universului în stadiile sale inițiale este, în multe privințe, foarte bine înțeleasă (vezi sfârșitul paragrafului 4.5).

Dar în absența ideilor teoretice de care avem nevoie, judecățile cu privire la existența conștienței rămân, deocamdată, ipoteze. Pentru a formula propriile mele ipoteze în această problemă, să spunem că am puternica convingere că, pe această planetă, existența conștiinței *nu* se limitează la ființele umane. În unul dintre cele mai emoționante programe de televiziune ale lui David Attenborough [3] era un episod care aproape că te convingea că elefanții, de exemplu, nu numai că simt emoții puternice, dar aceste emoții nu sunt foarte îndepărtate de cele care induc credințe religioase în oameni. Conducătorul unei turme de elefanți — o femelă a cărei soră murise cu cinci ani în urmă — și-a condus turma într-un lung ocol la locul în care sora ei murise și, odată ajunși acolo, a cules cu mare atenție craniul din grămada de oase, trecându-l celorlalți elefanți care l-au purtat dintr-o trompă în alta cu gesturi de tandrețe. Că elefanții posedă înțelegere a fost demonstrat convingător — chiar dacă oribil — de un alt program de televiziune [4]. Filmul făcut din elicopter în timpul unei vânători, a arătat clar teroarea care i-a cuprins pe elefanți când au înțeles că urmează măcelărirea întregii turme.

Există interesante dovezi privind conștiența (și autoreflectarea) la primate, și, personal, n-am nici o îndoială că fenomenul conștiinței este o trăsătură a vieții animale chiar mai 'puțin evoluate' decât atât. De exemplu, într-o emisiune de televiziune [5] referitoare la agilitatea, hotărârea și uluitoarele resurse de care dispun unele veverițe, am fost complet uimit de o secvență în care o veveriță realiza că mușcând sârma pe care se cățăra va putea dezlega containerul cu nuci suspendat la o oarecare distanță. E greu de văzut cum această 'idee' a putut să-i vină instinctiv, sau ca parte a experiențelor precedente. Pentru a aprecia consecințele pozitive ale acțiunii sale, veverița ar fi trebuit să aibă o înțelegere rudimentară a *topologiei* implicate în situația dată (vezi paragraful 1.19). Mi se pare că a fost vorba pur și simplu de *imaginație* din partea veverței — ceea ce cu siguranță pretinde conștiență!

Pare, de asemenea, indubitabil că problema care se pune este existența unor

grade de conștiință; nu este vorba doar de 'existența' sau 'inexistența' conștiinței ca atare. Chiar și în propria mea experiență pot să depun mărturie pentru momente în care am avut senzația că este prezentă într-un grad mai mare sau mai mic (ca, de exemplu, în cazul stării de vis, în care gradul de prezență al conștiinței pare mult mai mic decât în starea de veghe).

Cât de departe putem, atunci, merge 'în jos'? E un teren vast pentru opinii de toate felurile. În ce mă privește, am oarecare dificultăți în a crede că insectele au o cantitate prea mare din această calitate, mai ales de când am văzut un alt film documentar, în care o insectă o mânca pe alta cu o voracitate remarcabilă, aparent fără să-și dea seama că tocmai este mâncată, la rândul ei, de o a treia insectă. Cu toate acestea, așa cum am remarcat în paragraful 1.15, nu pot să nu recunosc că tipul de comportamente de care este în stare o furnică este extraordinar de complex și de subtil. Trebuie oare să credem că sistemele de control ale furnicilor, atât de eficiente sunt fără legătură cu principiul, oricare ar fi el, care ne dă nouă calitatea înțelegerii? Celulele lor neuronale de control au propriul lor citoschelet, iar dacă aceste citoschelete conțin microtubule care sunt capabile să reprezinte un loc de existență pentru acele stări cuantice coerente pe care le-am așezat la originea propriei noastre conștiințe, n-ar putea să fie vorba, și în cazul furnicilor, de aceeași calitate? Dacă microtubulele din creierul nostru posedă enorma sofisticare necesară pentru menținerea activității de coerență cuantică colectivă, atunci este greu de imaginat cum anume a putut dezvolta selecția naturală această facilitate doar pentru noi și unii dintre verii noștri multicelulari. Stările de coerență cuantică trebuie să existe și în animalele unicelulare eucariote, poate, cel mult, cu valoare și funcție diferită.

Coerența cuantică la scară mare *nu* implică, în sine, conștiința — altfel superconductorii ar fi conștienți! Și totuși este foarte posibil ca o astfel de coerență să fie o *parte* din ingredientele de care este nevoie pentru a fi conștient. În creierele noastre există un foarte mare grad de organizare, și cum conștiința pare să fie o trăsătură *globală* a gândirii noastre, pare că trebuie să căutăm un tip de coerență la scară mult mai mare decât la nivelul unui singur microtubul sau al unui microschelet. Ar trebui să existe o non-separabilitate semnificativă între stările din citoscheletele separate a unui mare număr de neuroni diferiți, astfel încât arii largi din creier să fie implicate într-un fel de stare cuantică colectivă. Însă este nevoie de mai mult decât atât. Pentru ca un anumit tip de non-computabilitate să fie

implicată şi susţin că asta este o parte esenţială a conştiinţei ar fi necesar ca sistemul să facă uz de aspectele non-aleatoare (non-computabile) ale OR. Propunerea specifică pe care am înaintat-o în paragraful 6.12 ne dă cel puţin o idee asupra dimensiunilor la care acţiunea OR începe să aibă importanţă.

Astfel, pe baza consideraţiilor din această carte, am putea într-un anumit fel anticipa un mod de a formula o ipoteză asupra nivelului la care conştiinţa conştientă ar putea începe să se manifeste.

Procesele care pot fi descrise în mod adecvat cu ajutorul fizicii computaţionale (sau aleatoare) nu vor implica conştiinţă, conform acestui punct de vedere. Pe de altă parte, chiar implicarea esenţială într-o acţiune non-computaţională OR precisă nu va implica, în sine, în mod necesar, prezenţa conştiinţei chiar dacă este o *precondiţie* pentru existenţa conştiinţei. Evident, nu avem de-a face cu un criteriu bine definit, însă e tot ce pot face pentru moment. Să vedem cât de departe putem ajunge cu acesta.

Voi încerca să dezvolt imaginea care rezultă, bazându-mă pe sugestiile din paragraful 6.12 privind nivelul la care ar trebui să apară frontiera dintre clasic şi cuantic precum şi pe speculaţiile biologice din paragrafele 7.5-7.7 conform cărora ar trebui să găsim că această frontieră este relevantă pentru interfaţa intern-extern a sistemului de microtubule dintr-o celulă, sau dintr-un sistem de celule. O idee adiţională esenţială este că dacă reducţia vectorului de stare apare doar din cauza unei părţi prea mari din mediu care devine non-separabilă de sistemul considerat, OR apare în mod efectiv ca un proces aleator pentru care argumentele FAPP standard (expuse în paragraful 6.6) sunt adecvate, iar OR se comportă exact ca **R**. Ceea ce avem nevoie este ca această reducţie să aibă loc exact în punctul în care intră în joc detaliile (necunoscute) non-computaţionale cu ale presupusei teorii OR. Cu toate că detaliile acestei teorii nu sunt cunoscute, putem, cel puţin în principiu, să obţinem o idee privind nivelul la care teoria ar putea începe să devină relevantă. Astfel, pentru ca aceste aspecte non-computaţionale ale OR să joace un rol important, ar fi necesar să fie menţinut un anumit fel de coerenţă cuantică până când cuplajul provoacă *exact* atâta mişcare de materie încât OR acţionează *înainte* ca mediul aleator să fie semnificativ implicat.

Descrierea pe care o propun pentru microtubule presupune că există „oscilaţii cuantice coerente" care au loc *în interiorul* tubulilor şi că acestea sunt slab cuplate cu o „activitate computaţională celulară de tip automat" care are loc în

în comutatoarele conformaționale ale dimerilor tubulinici de *pe* tuburi. Atâta timp cât oscilațiile cuantice rămân izolate, nivelul ar fi prea mic pentru ca OR să aibă loc. Totuși, cuplajul presupune ca tubulinele să fie de asemenea implicate în starea cuantică, iar la un anumit moment OR va avea loc. Ceea ce avem nevoie să se întâmple este ca OR să aibă loc *înainte* ca mediul microtubulelor să devină non-separabil cu starea cuantică, deoarece imediat ce asta se întâmplă, aspectele non-computaționale ale OR sunt pierdute, iar acțiunea este doar procesul aleatoriu **R**.

Astfel, ne putem întraba dacă într-o singură celulă (ca de exemplu parameciul, sau celula de ficat umană) cantitatea de activitate conformațională a tubulinelor poate implica suficientă mișcare de masă pentru satisfacerea criteriului din paragraful 6.12, astfel încât OR să acționeze, sau, în cazul contrar, acțiunea lui OR este întârziată până când mediul *este* perturbat — iar jocul (non-computațional) este pierdut. La prima vedere, pare să fie într-adevăr prea puțină mișcare de masă în activitatea conformațională din tubulină, și deci meciul pare pierdut la acest nivel. Însă pentru un număr mare de celule, situația apare mult mai promițătoare.

Poate că această imagine favorizează într-adevăr un punct de vedere în care cerința non-computațională pentru conștiință poate apărea doar pentru un număr mare de celule, cum ar fi în cazul unui creier (nu foarte mic). [6] Evident, trebuie să fim foarte precauți, în acest stadiu, cu concluziile definitive. Atât latura biologică cât și cea fizică a descrierii sunt prea neriguros formulate pentru a putea desprinde o concluzie clară privind implicațiile punctului de vedere pe care-l susțin. Chiar cu propunerile specifice pe care le-am prezentat, o cercetare ulterioară serioasă este necesară atât în ce privește biologia cât și fizica, implicate pentru a permite o ipoteză întemeiată privind locul în care conștiința își face apariția.

Există și alte probleme care trebuie luate în considerare. Cât anume din creier este implicat într-o stare conștientă? Foarte probabil, *nu* este vorba despre tot creierul. O mare parte din acțiunea acestuia pare să nu fie conștientă. Cerebelul (vezi paragraful 1.14), mai ales, pare să acționeze în întregime inconștient. El este răspunzător de controlul precis și delicat al acțiunilor noastre. Cerebelul este frecvent asociat unui calculator din cauza activității sale în întregime inconștiente. Ar fi desigur instructiv să știm ce diferențe esențiale pot exista în organizarea celulară și citoscheletală a cerebelului, în comparație cu scoarța cerebrală, care pare să aibă o relație mult mai directă cu fenomenul conștiinței. Ce este interesant este că dacă luăm în considerare simpla numărătoare a neuronilor, între cele două nu

există o diferenţă prea mare; în scoarţa cerebrală există poate de două ori mai mulţi neouroni decât în cerebel, însă, pe de altă parte, în cerebel se produc mai multe conexiuni sinaptice între celulele individuale (vezi paragraful 1.14, figura 1.6). În concluzie, cu siguranţă că ne trebuie un mecanism explicativ bazat pe ceva mult mai subtil decât simpla numărătoare a celulelor[1]

Poate că avem, de asemenea, câte ceva de câştigat dintr-un studiu al felului în care controlul inconştient exercitat de cerebel este 'învăţat' de la cel conştient. Ar putea foarte bine să existe o puternică similitudine în cazul procedurilor de învăţare ale cerebelului cu felul în care sunt 'antrenate' reţelele neuronale artificiale, în conformitate cu filozofia conexionistă. Dar chiar şi aşa, chiar dacă faptul că anumite acţiuni ale scoarţei cerebrale pot fi parţial înţelese în acest fel este *de asemenea* adevărat — lucru implicit în abordarea conexionistă a înţelegerii cortexului vizual — nu există nici un motiv să ne aşteptăm ca acelaşi lucru să fie adevărat pentru acele aspecte ale scoarţei cerebrale care implică conştienţă. Într-adevăr, aşa cum am argumentat în partea întâia, aceste funcţii cognitive evoluate în care este implicată conştiinţa, trebuie să fie foarte diferite de conexionism.

8.7. Trei lumi şi trei mistere

Ar trebui să aduc acum temele acestei cărţi împreună. Problema centrală pe care am urmărit-o de-a lungul întregii poveşti a fost felul în care fenomenul conştiinţei ar putea fi integrat într-o imagine ştiinţifică despre lume. Recunosc că nu am avut prea multe de spus despre problema conştiinţei în general. În loc de asta, m-am concentrat, în partea întâia, asupra unei calităţi mentale particulare,

[1] Ca profan în neuroanatomie nu pot să nu fiu şocat de faptul că există ceva ciudat (neexplicat?) în organizarea cerebrală care *nu* pare să fie aceeaşi pentru cerebel. Mulţi dintre senzorii şi nervii motori se încrucişează, astfel încât partea stângă a scoarţei cerebrale este 'răspunzătoare' de partea dreaptă a corpului şi invers. Şi nu numai atât, dar partea scoarţei răspunzătoare pentru vedere este în spate, pe când ochii sunt în faţă; partea corespunzătoare picioarelor este deasupra, în timp ce picioarele însele sunt în partea de jos a corpului, iar partea corespunzătoare auzului este diametral opusă faţă de ureche. Această inversare nu este o trăsătură universală a scoarţei, însă nu pot să nu cred că nu este vorba de un simplu accident în această dispunere, pentru că cerebelul nu este organizat astfel. Beneficiază conştiinţa în vreun fel din faptul că semnalele nervoase au de parcurs un drum mai lung?

înțelegerea conștientă, și în special înțelegerea matematică. Datorită acestei calități mentale speciale am fost în stare să formulez o ipoteză 'tare' conform căreia este în mod esențial *imposibil* ca o astfel de calitate să apară doar ca o trăsătură a activității computaționale sau să poată fi simulată computațional în mod adecvat. În plus, vreau să subliniez că nu este nimic special care să distingă înțelegerea matematică de orice alt tip de înțelegere. Concluzia acestei părți a fost că indiferent care este activitatea cerebrală responsabilă pentru nașterea conștiinței, ea trebuie să depindă de o fizică situată dincolo de simularea computațională. Partea a doua reprezintă o încercare de a găsi, în interiorul științei actuale, o acțiune fizică relevantă care ne-ar putea duce dincolo de limitările calculului. Pentru a pune împreună chestiunile cu care m-am confruntat, voi reformula lucrurile în funcție de trei lumi diferite și a relațiilor dintre ele. Aceste lumi sunt oarecum înrudite cu cele ale lui Popper (vezi Popper și Eccles, 1977) însă formularea mea va fi ușor diferită.

Lumea pe care o cunoștem cel mai direct este *lumea percepțiilor noastre conștiente*. Și tot aceasta este lumea despre care știm cel mai puțin în termeni științifici preciși. Este lumea care conține fericirea și durerea și percepția culorilor. Conține cele mai vechi amintiri din copilărie și frica noastră de moarte. Conține iubire, înțelegere și cunoașterea multor lucruri, așa cum conține de asemenea ignoranță și dorința de răzbunare. Este o lume care conține imagini mentale de mese și scaune și în care mirosuri, sunete și senzații de toate tipurile sunt amestecate cu gândurile și voința noastră de a acționa.

Dar mai există alte două lumi despre existența cărora știm câte ceva — mai puțin direct decât despre lumea percepțiilor noastre. Una dintre aceste lumi este ceea ce numim *lumea fizică*. Conține scaunele și mesele actuale, televizoare și automobile, ființe umane, creiere umane și acțiunile neuronilor. În această lume există soarele, luna și stelele, norii, uraganele, pietrele, florile și fluturii. La un nivel mai subtil există moleculele și atomii, electronii și fotonii sau spațiu-timpul. Tot această lume conține citoscheletele și tubulinele sau superconductorii. Nu este deloc clar dacă lumea percepțiilor nastre conștiente ar trebui să aibă vreo legătură cu lumea fizică, dar, cel puțin aparent, ele ni se înfățișează oarecum legate.

Există și o altă lume, a cărei existență e ceva mai dificil de acceptat: este *lumea platonică a formelor matematice*. Acolo găsim numerele naturale 0, 1, 2, 3...., sau algebra numerelor complexe. Tot acolo se află teorema lui Lagrange

conform căreia orice număr natural este suma a patru pătrate, geometria lui Euclid sau teorema lui Pitagora. În aceeași lume platonică există propoziția conform căreia pentru orice pereche a, b de numere naturale avem $a \times b = b \times a$ dar și faptul că acest rezultat nu este adevărat pentru anumite tipuri de 'numere' (ca în cazul produsului Grassmann la care ne-am referit în paragraful 5.15). Aceeași lume conține și alte geometrii în afara celei euclidiene, geometrii în care teorema lui Pitagora nu mai este adevărată. Conține numere infinite și necalculabile, ordinale recursive și non-recursive, mașini Turing care nu se opresc niciodată și mașini oracol. În ea se găsesc multe clase de probleme matematice insolvabile prin calcul. Și tot în această lume se află ecuațiile Maxwell pentru câmpul electromagnetic și ecuațiile gravitaționale ale lui Einstein, împreună cu un număr infinit de spații-timp care le satisfac (fie ele realiste sau nu din punct de vedere fizic). Aici avem simulări matematice ale scaunelor și meselor, așa cum apar ele în *realități virtuale*, ca și simulări ale găurilor negre sau ale uraganelor.

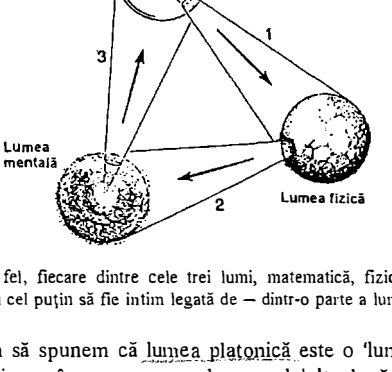


Fig. 8.1. Într-un anume fel, fiecare dintre cele trei lumi, matematică, fizică și mentală pare să 'emeargă' misterios sau cel puțin să fie intim legată de — dintr-o parte a lumii precedente.

Ce drept avem să spunem că lumea platonică este o 'lume' care 'există' în mod actual, în același sens în care spunem despre celelalte două lumi că există? Ea ar putea părea, desigur, doar o cutie cu concepte abstracte în care umblă matematicienii din când în când. Și totuși, existența ei rezidă în natura profundă, universală și atemporală a acestor concepte și în faptul că legile lor sunt independente de cei care le descoperă. Cutia asta nu a fost creația noastră. Numerele naturale au existat cu mult înaintea ființelor umane, sau a oricărei alte

creaturi de pe Pământ şi vor exista şi după ce orice altă formă de viaţă va fi demult dispărută. A fost dintotdeauna adevărat că un număr natural este suma a patru pătrate — n-a fost nevoie de Lagrange pentru ca acest rezultat să ia fiinţă. Numerele naturale atât de mari, încât se află în afara puterii de calcul a oricărui calculator, sunt tot suma a patru pătrate, chiar dacă nu avem nici cea mai mică şansă de a afla vreodată care sunt acestea. Va fi întotdeauna adevărat că nu există o procedură generală pentru a decide dacă acţiunea unei maşini Turing se opreşte vreodată şi a fost dintotdeauna adevărat, cu mult înainte de Turing însuşi.

Cu toate acestea, s-ar putea argumenta că natura absolută a adevărului matematic nu este un argument pentru a atribui o 'existenţă' reală conceptelor şi adevărurilor matematice. Aud uneori că platonismul matematic este desuet. E adevărat că Platon însuşi a murit cu 2340 de ani în urmă, însă ăsta nu e un motiv! O obiecţie ceva mai serioasă este cea legată de dificultăţile pe care le au filozofii atunci când este vorba ca o lume cu totul abstractă să afecteze lumea reală. Această profundă problemă este parte constitutivă a unuia dintre misterele despre care vom vorbi în curând. De fapt, realitatea conceptelor matematice este o idee mult mai naturală pentru un matematician decât pentru aceia care n-au avut norocul să-şi petreacă timpul explorând bogăţiile acestei lumi. Totuşi, pentru moment, nu este necesar să-mi acceptaţi punctul de vedere privind realitatea lumii matematice. Luaţi pur şi simplu 'lumea matematică a formelor platonice' ca pe o figură de stil, dacă vreţi; va fi o expresie de mare folos şi de mare însemnătate în cele ce urmează. Atunci când vom explora cele trei mistere care leagă *lumile* veţi vedea, probabil, că felul în care mi-am ales expresiile nu e întâmplător.

Dar ce sunt aceste mistere? Am încercat să le ilustrez în figura 8.1. În primul rând, este misterul privind faptul că aceste legi matematice profunde şi precise joacă un rol de o asemenea importanţă în comportamentul lumii fizice. Am putea spune chiar că, într-un anumit fel, lumea realităţii fizice emerge din lumea platonică a matematicii. Acest lucru este reprezentat prin săgeata înspre dreapta jos, care uneşte lumea platonică cu lumea fizică. Al doilea mister priveşte felul în care fiinţele capabile de percepţie pot apărea în lumea fizică. Cum se întâmplă că aceste obiecte materiale cu organizare complexă pot dezvolta în mod misterios entităţi mentale din substanţa materială a lumii fizice? Acest al doilea mister este reprezentat în figura 8.1 prin săgeata îndreptată în jos care uneşte lumea fizică cu cea mentală. Şi, în sfârţit, există misterul felului în care mentalul pare să fie în

stare să 'creeze' concepte matematice dintr-un fel de model mental. Aceste unelte aparent vagi, discutabile şi deseori nepotrivite cu care lumea noastră mentală pare înzestrată, par cu toate acestea să fie în stare, într-un mod misterios, (cel puţin atunci când sunt folosite la capacităţile maxime) să descrie forme matematice abstracte şi să ne permită astfel intrarea, cu ajutorul înţelegerii, în lumea platonică matematică.

Platon însuşi s-a preocupat mult de prima din aceste săgeţi (şi de asemenea, în felul său special, cu a treia), şi a avut grijă să distingă între forma matematică perfectă şi simpla ei *umbră* în lumea fizică. Astfel, un triunghi matematic (unul euclidian, aşa cum trebuie să avem grijă să specificăm în ziua de azi) va avea suma unghiurilor egală cu suma a două unghiuri drepte, pe când pentru unul făcut din lemn această sumă se va apropia mai mult sau mai puţin de cazul ideal, fără să fie vreodată în mod precis egală cu suma a două unghiuri drepte. Platon a descris aceste idei printr-o parabolă. El şi-a imaginat nişte cetăţeni închişi într-o peşteră, legaţi cu lanţuri în aşa fel încât să nu vadă decât peretele din fundul peşterii. Pe acest perete ei pot vedea umbrele unor forme perfecte care sunt purtate în spatele lor în lumina unui foc. Tot ce pot vedea sunt umbre imperfecte, distorsionate de lumina tremurândă a focului. Formele perfecte reprezintă obiectele matematice — umbrele, lumea *realităţii fizice*.

Din timpul lui Platon, rolul matematicii în structura perceptibilă şi în comportarea actuală a lumii a crescut enorm. Cunoscutul fizician Eugene Wigner a ţinut ân 1960 o celebră conferinţă cu titlul „Nerezonabila eficienţă a matematicii în ştiinţele fizice" în care a exprimat câte ceva din uimitoarea precizie şi subtila aplicabilitate a matematicilor sofisticate pe care fizicienii le găsesc continuu în descrierile lor asupra realităţii.

Pentru mine, cel mai impresionant exemplu este teoria relativităţii generalizate. Auzim adeseori că fiziceinii nu fac altceva decât să observe regularităţi în care conceptele matematice se întâmplă să se aplice cu succes comportamentului fizic. În consecinţă, se pretinde că fizicienii tind să se ocupe cu predilecţie de acele domenii în care descrierile matematice sunt eficiente, astfel încât nu este nici un mister în faptul că matematica poate fi aplicată în fizică. Un astfel de punct de vedere este însă cu totul deplasat dacă în discuţie este teoria lui Einstein; pur şi simplu nu este în stare să ofere o explicaţie pentru profunda unitate pe care teoria relativităţii o pune în evidenţă între matematică şi lumea fizică.

Atunci când teoria relativității generale a fost formulată, nu era decloc nevoie de ea din considerente observaționale. Teoria gravitațională a lui Newton rezistase timp de 250 de ani și dobândise în acest timp o remarcabilă precizie, de ordinul unu la zece milioane (deja o justificare impresionantă pentru a lua în serios nivelul matematic profund al realității fizice). Desigur, fusese observată o anomalie în mișcarea planetei Mercur, însă aceasta nu era un motiv suficient pentru a abandona schema newtoniană. Și totuși, Einstein a realizat, pe temeiuri fizice profunde, că am putea ajunge mai departe schimbând chiar cadrul teoriei gravitației. În cei câțiva ani după apariția teoriei lui Einstein s-au descoperit doar puține efecte care s-o susțină, iar creșterea gradului de precizie în raport cu teoria newtoniană era nesemnificativă. Cu toate acestea, acum, la aproape 80 de ani după elaborarea teoriei, precizia ei a crescut de aproximativ zece milioane de ori. Einstein nu se mulțumea să 'observe regularități' în comportamentul obiectelor fizice. El descoperea o substructură matematică profundă care era deja ascunsă în chiar acțiunile universului. Mai mult, el nu s-a limitat să caute în cele deja cunoscute un fenomen fizic care ar putea fundamenta mai bine o bună teorie. El a găsit aceste relații matematice precise în chiar structura spațiu-timpului — cea mai generală dintre noțiunile fizice existente.

În celelalte teorii de succes ale proceselor fizice fundamentale a existat întotdeauna o structură matematică subiacentă care s-a dovedit întotdeauna nu numai foarte precisă, dar și foarte sofisticată matematic. (Pentru cei care ar putea să-și imagineze că 'răsturnarea' ideilor fizicii precedente s-a datorat faptului că nu erau adecvate, mă simt obligat să precizez că lucrurile *nu* stau astfel. Ideile mai vechi, atunci când sunt bune, cum au fost cele ale lui Galilei și Newton, supraviețuiesc, și își au locul lor în noua schemă.) Mai mult, matematica însăși se inspiră din datele subtile și neașteptate pe care le oferă comportamentul detaliat al Naturii. Teoria cuantică — a cărei relație cu subtilitățile matematice este, sper, evidentă chiar și numai din sumara trecere în revistă a acestor pagini — ca și relativitatea generalizată și ecuațiile electromagnetice ale lui Maxwell, au furnizat o enormă capacitate de stimulare a progresului matematicii. Acest lucru este însă valabil nu nuami pentru teoriile recente ale fizicii. A fost cel puțin la fel de adevărat pentru cele mai vechi, că mecanica newtoniană (care a furnizat calculul diferențial și integral) sau analiza greacă a structurii spațiului (care ne-a dat chiar noțiunea de geometrie). Extraordinara acuratețe a matematicii în interiorul

comportamentelor fizice a fost adesea subliniată (precizia de până la a 11-a sau a 12-a zecimală din electrodinamica cuantică). Însă în misterul de care este vorba e mult mai mult decât atât. Este vorba, în primul rând, de remarcabila adâncime, subtilitate şi *fecunditate matematică* în conceptele latente în interiorul fizicii. Ori acesta este un lucru nefamiliar celor mai mulţi dintre oameni, dacă nu cumva se ocupă cu matematica.

Ar trebui să subliniem că această fecunditate matematică, furnizând un stimul valoros pentru activităţile actuale ale matematicianului, nu este doar o problemă de modă matematică (deşi moda joacă de asemenea un rol). Idei care au fost dezvoltate cu singurul scop de a adânci înţelegerea noastră privind acţiunile din lumea fizică s-au dovedit de multe ori interesante pentru formularea şi rezolvarea unor probleme matematice care erau de multă vreme privite ca obiecte de interes pentru cu totul alte motive. Unul dintre cele mai uimitoare exemple recente a fost folosirea teoriilor de tip Yang-Mills (care au fost dezvoltate de fizicieni pentru explicarea matematică a interacţiunilor dintre particulele elementare) de către Simon Donaldson, la Oxford, pentru obţinerea unor proprietăţi cu totul neaşteptate ale varietăţilor patru-dimensionale [8]. Astfel de proprietăţi matematice, cu nimic anticipate de oameni înainte ca abordările potrivite să fie imaginate, au avut o existenţă atemporală în lumea platonică, ca adevăruri imuabile care aşteaptă să fie descoperite.

Sper că v-am convins că între lumea platonică a matematicii şi lumea obiectelor fizice există o legătură reală şi strânsă — chiar dacă profund misterioasă. Sper, de asemenea, că prezenţa acestei indiscutabile legături îi va ajuta pe sceptici să ia ceva mai în serios, decât erau dispuşi până acum, lumea platonică *ca* 'lume' S-ar putea, de altfel, merge chiar mai departe decât am făcut-o până acum. Poate că şi celorlalte concepte abstracte, nematematice, li s-ar putea atribui o realitate platonică. Platon însuşi atribuia realitate conceptelor de *bine* şi de *frumos* (vezi paragraful 8.3). În ce mă priveşte, nu sunt împotriva unei asemenea posibilităţi; doar că ea nu joacă nici un rol în problemele de care mă ocup deocamdată. Chestiunile legate de etică, morală şi estetică nu au importanţă pentru discuţia prezentă, dar acesta nu este un motiv să le considerăm 'ireale'. Evident, sunt nenumărate lucruri importante de discutat în acest punct, însă ele nu constituie obiectul cărţii de faţă [9].

De asemenea, nu m-a interesat în mod special, în această carte, misterul

legăturii dintre lumea platonică şi lumea fizică. Am fost în schimb puternic preocupat de celelalte două, care sunt chiar mai puţin înţelese. În partea întâia am discutat chestiuni legate de cea de-a treia săgeată din figură: misterul percepţiilor noastre proprii a adevărului matematic, adică, a modului în care, prin contemplare mentală, părem să avem capacitatea de a face să apară formele matematice platonice. E ca şi cum formele perfecte ar fi simple umbre ale gândurilor noastre imperfecte. A vedea lumea platonică în acest mod — ca un simplu produs al lumii noastre mentale — ar fi cu totul contrar concepţiilor lui Platon. Pentru el, lumea formelor perfecte este primordială, fiind atemporală şi independentă. În concepţia platonică, cea de-a treia săgeată pe care am desenat-o în figura 8.1 ar trebui să fie îndreptată în jos în loc să arate în sus: din lumea formelor perfecte în cea a mentalului. A gândi lumea matematică ca pe un produs al modului nostru de gândire înseamnă să adoptăm o poziţie *kantiană* şi nu pe cea platonică.

De asemenea, unii ar putea susţine că ar trebui să inversez şi direcţiile celorlalte săgeţi. Probabil că episcopul Berkeley ar fi preferat ca a *doua* săgeată să arate dinspre lumea mentală înspre lumea experienţei fizice, o 'realitate fizică' fiind, în acest caz, doar o simplă umbră a existenţei noastre mentale. Există şi alţii ('nominaliştii') care ar argumenta în favoarea inversării *primei* săgeţi, lumea matematicii fiind în cazul lor doar o reflexie a lumii fizice. Opiniile mele, după cum vă este cred evident după citirea acestei cărţi, se îndreaptă împotriva inversării sensului primelor două săgeţi. În acelaşi timp, cred că este de asemenea evident că sunt, într-o anumită măsură nemulţumit de aparenta orientare 'kantiană' a celei de-a treia! Pentru mine, lumea formelor matematice perfecte este primordială (după cum credea şi Platon) — existenţa ei fiind aproape o necesitate logică — pe când celelalte două lumi sunt doar umbrele ei.

Luând însă în considerare toate aceste opinii, voi propune să priviţi săgeţile într-o lumină oarecum diferită. Chestiunea esenţiază în ce le priveşte nu este atât direcţia lor, cât faptul că în fiecare caz ele reprezintă o corespondenţă între o *mică* regiune a unei lumi care cuprinde *toată* lumea următoare. În ce priveşte prima săgeată, a fost adesea remarcat că cea mai mare parte a lumii matematice (judecată din perspectiva activităţii matematicianului) are prea puţin de-a face cu comportamentul fizic al obiectelor. Astfel, doar o mică parte a lumii platonice poate sta la baza structurii universului fizic. La fel, a doua săgeată exprimă faptul că existenţa noastră mentală emerge doar dintr-o minusculă porţiune a lumii fizice

— un domeniu organizat într-un mod foarte precis, care permite apariţia conştiinţei, ca, de exemplu, creierul uman. În acelaşi fel, a treia săgeată se referă la faptul că doar o mică parte din activitatea noastră mentală este preocupată de chestiuni abstracte şi atemporale ca matematica. În cea mai mare parte, viţile noastre sunt pline de cu totul alte lucruri!

Există şi un aparent aspect paradoxal al acestor corespondenţe, în care fiecare lume pare să 'emeargă' dintr-o minusculă regiune a lumii care o precede. Am desenat figura 8.1 astfel încât să subliniez acest paradox. Totuşi, privind săgeţile doar ca expresii ale diferitelor corespondenţe şi nu ca expresia vreunei emergenţe actuale, încerc să nu intru în dezbateri asupra întrebării privind care dintre aceste lumi trebuie privită ca primară, secundară şi terţiară.

Şi totuşi, chiar şi aşa, figura 8.1 exprimă alt aspect al propriilor mele opinii, ori poate prejudecăţi. Am reprezentat lucrurile ca şi când fiecare lume întreagă este într-adevăr reflectată într-o mică porţiune a lumii precedente. Poate că presupoziţiile mele sunt greşite. Poate că există aspecte ale lumii fizice care *nu pot fi descrise* în termeni matematici precişi; poate că viaţa mentală *nu* este înrădăcinată în structurile fizice (creiere, de exemplu). Poate există adevăruri matematice care rămân, în principiu, *inaccesibile* raţiunii umane. Pentru a cuprinde oricare dintre aceste posibile alternative, figura 8.1 ar trebui redesenată, astfel încât să permită unora dintre aceste lumi să se extindă dincolo de aria sub-întinsă de săgeata precedentă.

În partea întâi, am tratat pe larg unele dintre implicaţiile faimoasei teoreme de incompletitudine a lui Gödel. Unii cititori ar putea crede că teorema lui Gödel ne spune că trebuie să existe părţi ale lumii adevărurilor matematice care se află, în principiu, în afara înţelegerii umane. Sper că argumentele mele v-au arătat că nu acesta este cazul [10]. Propoziţiile matematice specifice furnizate de argumentul lui Gödel sunt accesibile înţelegerii umane — cu condiţia să fie construite din sisteme matematice formale care au fost deja acceptate ca metode valide de a aserta adevărul matematic. Teoremele lui Gödel nu argumentează în favoarea existenţei unui adevăr matematic inccesibil, ci în favoarea faptului că înţelegerea şi intuiţia umană se află dincolo de argumente formale şi proceduri calculabile. Mai mult, argumentează puternic în favoarea existenţei lumii matematice platonice. Adevărul matematic nu este determinat arbitrar de regulile vreunui sistem formal construit de om, ci are un caracter absolut şi se găseşte

dincolo de orice sistem de reguli specificabile. De altfel, sprijinul în favoarea unui punct de vedere platonic (ca opus celui formalist) a fost una dintre motivațiile inițiale ale lui Gödel. Pe de altă parte, argumentele din teorema lui Gödel servesc la ilustrarea naturii cu adevărat misterioase a percepțiilor noastre matematice. Nu este destul să 'calculăm' pentru a forma percepții. La mijloc este ceva cu totul diferit: ceva care ar fi cu desăvârșire imposibil în afara conștiinței conștiente la care se reduce, până la urmă, lumea percepțiilor.

Partea a doua a cărții s-a ocupat în special cu chestiuni legate de a doua săgeată (deși săgețile nu pot fi separate) — datorită căreia lumea fizică concretă poate invoca *umbra* pe care o numim conștiință. Cum se poate ca fenomenul conștiinței să apară din astfel de ingredienți nepromițători ca materia, spațiul și timpul? Nu am ajuns la un răspuns, dar sper că am putut cel puțin să vă conving că problema este *misterioasă* în sine, cel puțin tot atât cât cadrul spațio-temporal în care operează fizica actuală. Pur și simplu nu știm care este natura materiei și legile care o guvernează suficient de bine pentru a înțelege ce fel de organizare este cea care dă naștere ființelor conștiente. Mai mult, cu cât examinăm mai adânc natura materiei, cu atât apare ea mai eluzivă, misterioasă și matematică. Am putea întreba: ce *este* materia conform celor mai bune instrumente științifice? Răspunsul se prezintă sub forma matematicii — nu atât cea a unui sistem de ecuații, cât a unor concepte matematice subtile pentru înțelegerea cărora e nevoie de mult timp și de multă bătaie de cap.

Dacă relativitatea generalizată ne-a arătat că până și noțiunile de spațiu și timp au suferit o profundă modificare, devenind mai misterioase și mai matematice, mecanica cuantică a arătat, chiar într-un grad mai mare, cum conceptul nostru de *materie* a avut o soartă similară. Nu numai materia, dar până și conceptul nostru de actualitate a avut de suferit o profundă schimbare. Cum poate simpla *posibilitate* contrafactuală ca ceva să se întâmple să poată avea o influență decisivă asupra a ceea ce *se întâmplă* în mod actual? Există ceva în misterul mecanicii cuantice care *pare*, cel puțin, mult mai aproape decât fizica clasică de tipul de mister necesar pentru a introduce mentalul în lumea realității fizice. Nu am nici o îndoială că atunci când vom avea teorii mai profunde, locul minții în relație cu fizica nu va mai părea atât de ieșit din comun ca astăzi.

În paragrafele 7.7 și 8.6 am încercat să văd care ar fi circumstanțele fizice potrivite apariției conștiinței. Trebuie totuși să subliniez că nu consider conștiința

doar ca pe cantitatea potrivită de mișcare de masă corespunzătoare unei teorii OR a frontierei dintre clasic și cuantic. Așa cum sper că am explicat, astfel de lucruri nu fac decât să furnizeze o deschidere spre existența unei acțiuni non-computaționale în interiorul fizicii de astăzi. Conștiința propriu-zisă implică conștiența asupra unei varietăți infinite de lucruri calitativ diferite — culoarea verde a unei frunze, parfumul unui trandafir, cântecul unei mierle, moliciunea catifelată a blănii pisicii, dar și trecerea timpului, stările emoționale, îngrijorarea, minunarea, încântarea în fața unei idei. Presupune speranțe, idealuri, intenții și voința care direcționează nenumărate mișcări diferite ale corpului pentru ca aceste intenții să fie realizate. Studiul neuroanatomiei, al dezordinilor mentale, psihiatria și psihologia ne-au spus destul de multe despre relația detaliată dintre natura fizică a creierului și condiția noastră mentală. Nici nu se pune problema să putem înțeleg astfel de lucruri doar în termenii unor cantități critice de masă în mișcare. Și totuși, fără o astfel de deschidere spre o nouă fizică, rămânem blocați în forma îngustă a fizicii computaționale, sau a unei fizici computaționale cu elemente aleatoare. În această formă strâmtă nu poate exista vreun rol științific pentru intenționalitate și experiență subiectivă. Rupând cu această fizică, deschidem cel puțin posibilitatea unui astfel de rol.

Mulți dintre cei care ar putea fi de acord cu aceste argumente ar putea spune că nu poate exista *nici* un fel de tratare științifică a conștiinței. Pentru cei care susțin acest punct de vedere nu pot să cer decât timp și răbdare pentru a vedea evoluția viitoare a științei. Cred că avem deja indicii, în interiorul mecanicii cuantice, că mentalul este ceva mai aproape de înțelegerea noastră decât a fost înainte — deși doar cu *puțin* mai aproape. Cred că atunci când *noile* dezvoltări fizice necesare vor vedea lumina zilei, toate aceste indicii vor deveni mult mai clare. Știința mai are un lung drum în față: de *asta* sunt foarte sigur!

Mai mult, însăși posibilitatea înțelegerii umane a acestor lucruri ne spune câte ceva în privința posibilităților pe care conștiința ni le oferă. Trebuie să admitem că sunt unii, ca Newton sau Einstein, sau Arhimede, Galilei, Maxwell sau Dirac — sau Darwin, Leonardo da Vinci, Rembrandt, Picasso, Bach, Mozart, sau Platon, sau cei ale căror minți au conceput *Iliada* sau *Hamlet* — care par să fi avut mai mult din această facultate, destul pentru a 'mirosi' direct adevărul, sau frumosul. Pe de altă parte, există lucrările Naturii — o unitate care se manifestă potențial în noi toți și este manifestă în facultățile de înțelegere conștientă și

sensibilitate, indiferent la ce nivel operează acestea. Fiecare din creierele noastre conştiente este făcut din ingrediente fizice subtile care ne permit cumva să beneficiem de pe urma organizării matematice profunde a universului nostru – astfel încât, fiecare dintre noi este capabil, în schimb, de un fel de acces direct, prin intermediul calităţii platonice a 'înţelegerii', la chiar felul în care universul nostru se comportă la cele mai diferite nivele ale sale.

Toate acestea sunt probleme profunde, iar noi ne găsim încă departe de o explicaţie. Aş spune că nu vom ajunge la răspunsuri clare câtă vreme nu vor intra în joc interacţiunile dintre *toate* cele trei lumi. Nici una dintre întrebări nu se poate rezolva izolat de celelalte. Am vorbit despre trei lumi şi misterele care le leagă. Fără îndoială, nu există trei lumi, ci doar *una*, a cărei adevărată natură nici măcar n-o întrezărim, deocamdată.

Note şi referinţe bibliografice

1. Vezi, de exemplu, Lisboa, (1992)
2. Această idee mi-a fost descrisă de Joel de Rosnay.
3. *Echo of the elephants*, (BBC, Ianuarie, 1993)
4. *If the rains don't come* (BBC, Septembrie, 1992)
5. *Daylight robbery* (BBC, August, 1993)
6. S-ar putea specula pe marginea absenţei centriolilor din neuroni (conform paragrafului 7.4). Citoscheletele altor tipuri de celule individuale par să aibă nevoie de centrioli ca „centre de control". Poate că citoscheletele neuronilor ţin de o autoritate mai globală!
7. Marr (1982), Brady (1993).
8. Donaldson (1983), Devlin(1988), cap.10 pentru o abordare care nu este tehnică.
9. Lucrarea „World 3" a lui Popper conţine constructe mentale cu o oarecare similitudine cu cele din lumea platonică matematică. Vezi Popper, Eccles, (1977). Totuşi, a treia lume nu este văzută ca atemporală şi independentă de noi înşine, nici ca o lume aflată la baza structurii însăşi a realităţii fizice. De aici, statutul ei diferit de lumea platonică de care vorbesc eu.
10. Mostowski (1957) subliniază în introducerea la cartea sa că argumentele ca cele ale lui Gödel nu se referă la întrebarea dacă pot exista chestiuni matematice *absolut* indecidabile. Problema trebuie privită ca fiind complet deschisă, deocamdată şi rămâne, ca celelalte două, o problemă de credinţă!

Epilog

Jessica şi tatăl ei ieşiră din peşteră. Se întuneca şi începeau să se vadă stelele. Jessica spuse:

„Tati, ştii, când mă uit la cer, îmi vine greu să cred că Pământul chiar se mişcă că se învârte în jurul lui cu toţi acei mii de kilometri pe oră chiar dacă cred că ştiu acum că *trebuie* să fie adevărat.

Făcu o pauză, uitându-se la cer pentru o vreme.

„Tati, spune-mi despre stele..."

Boolos, G. (1990). On seeing the truth of the Gödel sentence. *Behavioural and Brain Sciences*, **13** (4), 655

Bowie, G.L. (1982). Lucas' number is finally up. *J. of Philosophical Logic*, **11**, 279-285

Brady, M. (1983). Computational vision. In *The simulation of human intelligence* (ed. D. Broadbent), Blackwell, Oxford

Braginsky, V.B. (1977). The detection of gravitational waves and quantum non-disturbative measurements. In *Topics in theoretical and experimental gravitation physics* (ed. V. de Sabbata şi J. Weber),p. 105, Plenum, London

Broadbent, D. (1993). Comparison with human experiments. In The simulation of human intelligence (ed. D. Broadbent), Blackwell Oxford

Brown, H.R. (1993). Bell's other therem and its connection with nonlocality. Part I in *Bell's theorem and the foundations of physics* (ed. A. Van der Merwe şi F. Selleri), World Scientific, Singapore

Butterfield, J. (1990). Lucas revived? An undefended flank. *Behavioural and Brain Sciences,* **13(4)**.658

Castagnoli, G., Rasetti, M. şi Vincenti, A. (1992). Steady, simultaneous quantum computation: a paradigm for the investigation of nondeterministic and non-recursive computation. *Int. J. Mod. Phys.*, C, **3**, 661-689

Caudill, M. (1992). *In our own image. Building an artificial person.* Oxford University Press.

Chaitin, G.J. (1975). Randomness and mathematical proof. *Scientific American*, (May 1975), 47

Chalmers, D.J. (1990). Computing the thinktable. *Behavioural and Brain Sciences*, **13(4)**, 658

Chandrasekhar, S. (1987). *Truth and beauty. Aesthetics and motivations in science.* The University of Chicago Press

Chang, C.-L. and Lee, R.C.-T. (1987). Symbolic logic and mechanical theorem proving, 2nd edn (ed.1-a în 1973) Academic Press, New York

Chou, S.-C. (1988). Mechanical geometry theorem proving. Ridel

Christian, J.J. (1994). On definite events in a generally covariant quantum world. Unpublished preprint

Church, A. (1936). An unsolvable problem of elementary number theory. *Am. Jour. of Math.*, **58**, 345-363

Church, A. (1941). *The calculi of lambda-conversion.* Annals of Mathematics Studies, nr.6, Princeton University Press

Churchland, P.M. (1984). *Matter and consciousness.* Bradford Books, MIT Press, Cambridge, Massachusetts

Clauser, J.F. and Horne, M.A. (1974). Experimental consequences of objective local theories. *Phys. Rev.*, **D10**, 526-535

Clauser, J.F., Horne, M.A. şi Shimony, A. (1978). Bell's theorem: experimental tests and implications. *Rpts. on Prog. in Phys.*, **41**, 1881-1927

Cohen, P.C. (1966). *Set theory and the continuum hypothesis.* Benjamin, Menlo Park, CA

Conrad, M. (1990). Molecular computing. In *Advanced in computers* (ed. M.C. Yovits), Vol.31, Academic Press, London

Conrad, M. (1992). Molecular computing: the lock-key paradigm. *Computer* (November 1992), 11-20

Conrad, M. (1993). The fluctuon model of Force, Life, and computation: a constructive analysis. *Appl. Math. and Comp.*, **56**, 203-259

Craik, K. (1943). *The nature of explanation.* Cambridge University Press

Crick, F. (1994). *The astonishing hypothesis. The scientific search for the soul.* Charles Scribner's Sons, New York, and Maxwell Macmillan International

Crick, F. and Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, **2**, 263-275

Crick, F. and Koch, C. (1992). The problem of consciousness. *Sci. Amer.*, **267**, 110

Curl, R.F. and Smalley, R.E. (1991). Fullerenes. *Scientific American*, **265**, nr.4, 32-41

Cutland, N.J. (1980). *Computability. An introduction to recursive function theory.* Cambridge University Press

Davenport, H. (1952). *The higher arithmetic.* Hutchinson's University Library

Davies, P.C.W. (1974). *The physics of time asymmetry.* Surrey University Press, Belfast

Davies, P.C.W. (1984). *Quantum mechanics.* Routledge, London

Davis, M. (ed.) (1965). *The undecidable basic papers on undecidable propositions, unsolvable problems and computable functions.* Raven Press, Hewlett, New York

Davis, M. (1978). What is a computation? In *Mathematics today; twelve informal essays* (ed. L.A. Steen), Springer-Verlag, New York

Davis, M. (1990). Is mathematical insight algorithmic? *Behavioural and Brain Sciences,* **13**(4), 659

Davis, M. (1993). How subtle is Gödel's theorem? *Behavioural and Brain Sciences,* 16, 611-612

Davis, M. and Hersch, R. (1975). Hilbert's tenth problem. *Scientific American* (Nov 1973), 84

Davis, M. and Hersch, R. (1982). *The mathematical experience.* Harvester Press

de Beauregard, C. and Werbos, P. (1989). In *Bell's theorem, quantum theory and conceptions of the universe* (ed. M. Kafatos). Kluwer, Dordrecht

de Broglie, L. (1956). *Tentative d'interpretation causale et nonlineaire de la mechanique*

ondulatoire. Gautier-Villars, Paris

Deeke, L., Grötzinger, B. şi Kornhuber, H.H. (1976). Voluntary finger movements in man: cerebral potentials and theory. *Biol. Cybernetics,* **23**,99

del Giudice, E., Doglia, S. şi Milani, M. (1983). Self-focusing and ponderomotive forces of coherent electric waves a mechanism for cytoskeleton formation and dynamics. In *Coherent excitations in biological systems* (ed. H. Fröhlich and F. Kremer), Springer-Verlag, Berlin

Dennett, D. (1990). Betting your life on an algorithm. *Behavioural and Brain Sciences*, **13(4)**,,660

Dennett, D.C. (1991). *Consciousness explained.* Little, Brown and Company

d'Espagnat, B. (1989). *Conceptual foundations of quantum mechanics*, 2nd edn. Addison-Wesley, Reading, Massachusetts

Deutsch, D. (1985). Quantum theory, the Church-Turing principle and the universal quantum computer. *Proc. Roy. Soc. (Lond.),* **A400**, 97-117

Deutsch, D. (1989). Quantum computational networks. *Proc. Roy. Soc. (Lond.),* **A425**, 73-90

Deutsch, D. (1991). Quantum mechanics near closed time-like lines. *Phys. Rev.,* **D44**, 3197-3217

Deutsch, D. (1992). Quantum computation. *Phys. World*, **5**, 57-61

Deutsch, D. and Ekert, A. (1993). Quantum communication moves into the unknown. *Phys. World,* **6**, 22-23

Deutsch, D. and Jozsa, R. (1992). Rapid solution of problems by quantum computation. *Proc. Roy. Soc. Lond.*, **A439**, 553-558

Devlin, K. (1988). *Mathematics: the New Golden Age.* Penguin Books, London

DeWitt, B.S. and Graham, R.D. (ed.) (1973). *The many-worlds interpretation of quantum mechanics.* Princeton University Press

Dicke, R.H. (1981). Interaction-free quantum measurements: a paradox? *Am. J. Phys.*, **49**, 925-930

Diósi, L. (1989). Models for universal reduction of macroscopic quantum fluctuations. *Phys. Rev.,* **A40**, 1165-1174

Diósi, L. (1992). Quantum measurement and gravity for each other. In *Quantum chaos, quantum measurement*; NATO AS1 Series C. Math. Phys. Sci. 357 (ed. P. Cvitanovic, I.C. Percival, A. Wirzba), Kluwer, Dordrecht

Dirac, P.A.M. (1947). *The principles of quantum mechanics*, 3rd edn. Oxford University Press

Dodd, A. (1991). Gödel, Penrose, and the possibility of AI. *Artificial Intelligence Review,* **5**

Donaldson, S.K. (1983). An application of gauge theory to four dimensional topology. *J. Diff. Geom.*, **18**, 279-315

Doyle, J. (1990). Perceptive questions about computation and cognition. *Behavioural and Brain Sciences*, **13(4)**, 661

Dreyfus, H.L. (1972). *What computers can't do.* Harper and Row, New York

Dummett, M. (1973). *Frege: philosophy of language.* Duckworth, London

Dustin, P. (1984). *Microtubules*, 2nd revised edn. Springer-Verlag, Berlin

Dryl, S. (1974). Behaviour and motor responses in paramecium. In *Paramecium - a current survey* (ed. W.J. Van Wagtendonk), 165-218, Elsevier, Amsterdam

Eccles, J.C. (1973). *The understanding of the brain.* McGraw-Hill, New York

Eccles, J.C. (1989). *Evolution of the brain: creation of the self.* Routledge, London

Eccles, J.C. (1992). Evolution of conciousness. *Proc. Natl. Acad. Sci.*, 89, 7320-7324

Eccles, J.C. (1994). *How the self controls its brain.* Springer-Verlag, Berlin

Eckert, R., Randall, D. şi Augustine, G. (1988). *Animal physiology. Mechanisms and adaptations*, Chapter 11, Freeman, New York

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M.,Kruse, W., Munk, M. şi Reitboeck, H.J. (1988). Coherent oscillations: a mechanism of feature linking in the visual cortex? *Biol. Cybern.*, **60**, 121-130

Edelman, G.M. (1976). Surface modulation and cell recognition on cell growth, *Science*, **192**, 218-226

Edelman, G.M. (1987). *Neural Darwinism, the theory of neuronal group selection.* Basic Books, New York

Edelman, G.M. (1988). *Topobiology, an introduction to molecular embriology.* Basic Books, New York

Edelman, G.M. (1989). *The remembered present. A biological theory of consciousness.* Basic Books, New York

Edelman, G.M. (1992). *Bright air, brilliant fire: on the matter of the mind.* Allen Lane, The Penguin Press, London

Einstein, A., Podolsky, P. şi Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? In *Quantum theory and measurement* (ed. J.A. Wheeler and W.H. Zurek), Princeton University Press, 1983. Original în *Phys. Rev.*, **47**, 777-780

Elitzur, A.C. and Vaidman, L. (1993). Quantum-mechanical interaction-free measurements. *Found. of Phys.*, **23**, 987-997

Elkies, Noam G. (1988). On $A^4+B^4+C^4 = D^4$. *Maths. of Computation*, **51**, (nr.184) 825-835

Everett, H. (1957). 'Relative State' formulation of quantum mechanics. In *Quantum theory*

and measurement (ed. J.A. Wheeler and W.H. Zurek), Princeton University Press 1983; original in *Rev. of Modern Physics,* **29**, 454-462

Feferman, S. (1988). Turing in the Land of O(z). In *The universal Turing machine: a half-century survey* (ed. R. Herken), Kammerer and Unverzagt, Hamburg

Feynman, R.P. (1948). Space-time approach to non-relativistic quantum mechanics. *Revs. Mod. Phys.*, **20**, 367-387

Feynman, R.P. (1982). Simulating physics with computers. *Int. J. Theor. Physics,* **21(6/7)**, 467-488

Feynman, R.P. (1985). Quantum mechanical computers. *Optics News*, Feb., 11-20

Feynman, R.P. (1986). Quantum mechanical computers. *Foundations of Physics*, **16(6)**, 507-531

Fodor, J.A. (1983). *The modularity of mind.* MIT Press, Cambridge, Massachusetts

Franks, N.P. and Lieb, W.R. (1982). Molecular mechanics of general anaesthesia. *Nature*, **300**, 487-493

Freedman, D.H. (1994). *Brainmarkers*. Simon and Schuster, New York

Freedman, S.J. and Clauser, J.F. (1972). Experimental test of local hidden-variable theories. In *Quantum theory and measurement* (ed. J.A. Wheeler and W.H. Zurek), Princeton University Press 1983; original in *Phys. Rev. Lett.*, **28**, 938-941

Frege, G. (1964). *The basic laws of arithmetic*, translated and edited with an introduction by Montgomery Firth. University of California Press, Berkeley

Frege, G. (1893). *Grundgesetze der Arithmetik, begriffschriftlich abelgeleitet*, Vol.1, H. Pohle, Jena

French, J.W. (1940). Trial and error learning in paramecium. *J. Exp. Psychol.*, **26**, 609-613

Fröhlich, H. (1968). Long range coherence and energy storage in biological systems. *Int. Jour. of Quantum Chem.*, **II**, 641-649

Fröhlich, H. (1970). Long-range coherence and the actions of enzymes. *Nature*, **228**, 1093

Fröhlich, H. (1975). The extraordinary dielectric properties of biological materials and the action of enzymes. *Proc. Natl. Acad. Sci.*, **72(11)**, 4211-4215

Fröhlich, H. (1984). General theory of coherent excitations on biological systems. In *Nonlinear electrodynamics in biologicalsystems* (ed. W.R. Adey and A.F. Lawrence), Plenum Press, New York

Fröhlich, H. (1986). Coherent excitations in active biological systems. In *Modern bioelectrochemistry* (ed. F. Gutmann and H. Keyzer), Plenum Press, New York

Fukui, K. and Asai, H. (1976). Spiral motion of paramecium caudatum in small capilary glass tube. *J. Protozool.*, **23**, 559-563

Gandy, R. (1988). The confluence of ideas in 1936. In *The universal Turing machine: a*

half-century survey (ed. R. Herken), Kammerer and Unverzagt, Hamburg

Gardner, M. (1965). *Mathematical magic show.* Alfred Knopf, New York and Random House, Toronto

Gardner, M. (1970). Mathematical games: the fantastic combinations of John Conway's new solitaire game 'Life' *Scientific American,* **223**, 120-123

Gardner, M. (1989). *Penrose tiles to trapdoor ciphers.* Freeman, New York

Gelber, B. (1958). Retention in paramecium aurelia. *J. Comp. Physiol. Psych.*, **51**, 110-115

Gelernter, D. (1994). *The muse in the machine.* The Free Press, Macmillan Inc., New York and Collier Macmillan, London

Gell-Mann, M. and Hartle, J.R. (1993). Classical equations for quantum systems. *Phys. Rev.*, **D47**, 3345-3382

Gernoth, K.A., Clark, J.W., Prater, J.S. şi Bohr, H. (1993). Neural network models of nuclear systematics. *Phys. Lett.*, **B300**, 1-7

Geroch, R. (1984). The Everett interpretation. *Nous,* **4** (special issue on the foundations of quantum mechanics), 617-633

Geroch, R. and Hartle, J.B. (1986). Computability and physical theories. *Found. Phys.*, **16**, 533

Ghirardi, G.C., Rimini, A. şi Weber, T. (1980). A general argument against superluminal transmission through the quantum mechanical measurement process. *Lett. Nuovo Chim.*, **27**, 293-298

Ghirardi, G.C., Rimini, A. şi Weber, T. (1986). Unified dynamics for microscopic and macroscopic systems. *Phys. Rev.*, **D34**, 470

Ghirardi, G.C., Grassi, R. şi Rimini, A. (1990a). Continuous-spontaneous reduction model involving gravity. *Phys. Rev.*, **A42**, 1057-1064

Ghirardi, G.C., Grassi, R. şi Pearle, P. (1990b). Relativistic dynamical reduction models: general framework and examples. *Foundations of Physics,* **20**, 1271-1316

Ghirardi, G.C., Grassi, R. şi Pearle, P. (1992). Comment on 'Explicit collapse and superluminal signals' *Phys. Lett.*, **A166**, 435-438

Ghirardi, G.C., Grassi, R. şi Pearle, P. (1993). Negotiating the tricky border between quantum and classical. *Physics Today,* **46**, 13

Gisin, N. (1989). Stochastic quantum dynamics and relativity. *Helv. Phys. Acta.*, **62**, 363-371

Gisin, N. and Percival, I.C. (1993). Stochastic wave equations versus parallel world components. *Phys. Lett.*, **A175**, 144-145

Gleick, J. (1987). *Chaos. Making a new science.* Penguin Books

Glymour, C. and Kelly, K. (1990). Why you'll never know whether Roger Penrose is a

computer. *Behavioural and Brain Sciences,* 13(4), 666

Gödel, K. (1931). Über formal unentscheidbare Sätze per Principia Mathematica and verwandter Systeme I. *Monatschefte für Mathematik and Physik,* **38**, 173-198

Gödel, K. (1940). *The consistency of the axiom of choice and of the generalized continuum-hypothesis with the axioms of the set theory.* Princeton University Press and Oxford University Press

Gödel, K. (1949). An example of a new type of cosmological solution of Einstein's field equations of gravitation. *Rev. of Mod. Phy.,* **21**, 447

Gödel, K. (1986). *Kurt Gödel, collected works,* Vol. I (publications 1938-1974) (ed. S. Feferman et al.),Oxford University Press

Gödel, K. (1990). *Kurt Gödel, collected works,* Vol. II (publications 1938-1974) (ed. S. Feferman et al.),Oxford University Press

Gödel, K. (1995). *Kurt Gödel, collected works,* Vol. III (ed. S. Feferman et al.),Oxford University Press

Golomb, S.W. (1965). *Polyominoes,* Scribner and Sons

Good, I.J. (1965). Speculations concerning the first ultra intelligent machine. *Advanced Computers,* **6**, 31-88

Good, I.J. (1967). Human and machine logic. *Brit. J. Philos. Sci.,* **18**, 144-147

Good, I.J. (1969). Gödel's theorem is a red herring. *Brit. J. Philos. Sci.,* **18**, 359-373

Graham, R.L. and Rothschild, B.L. (1971). Ramsey's theorem for n-parameter sets. *Trans. Amer. Math. Soc.,* **59**, 290

Grant, P.M. (1994). Another December revolution? *Nature,* **367**, 16

Gray, C.M. and Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl. Acad. Sci. USA,* **86**, 1689-1702

Grangier, P., Roger, G. şi Aspect, A. (1986). Experimental evidence for a photon anticorrelation effect on a beam splitter: a new light on single-photon interferences. *Europhysics Letters,* **1**, 173-179

Green, D.G. and Bossomaier, T. (ed)(1993). *Complex systems: from biology to computation.* IOS Press

Greenberger, D.M., Horne, M.A. şi Zeilinger, A. (1989). Going beyond Bell's theorem. In *Bell's theorem, quantum theory, and conceptions of the universe* (ed. M. Kafatos) p. 73-76, Kluwer Academic, Dordrecht, The Netherlands

Greenberger, D.M., Horne, M.A., Shimony, A. şi Zeilinger, A. (1990). Bell's theorem without inequalities. *Am. J. Phys.,* **58**, 1131-1143

Gregory, R.L. (1981). *Mind in science: a history of explanations in psychology and physics.* Weidenfeld and Nicholson Ltd. (also Penguin 1984)

Grey Walter, W. (1953). *The living brain.* Gerald Duckworth and Co. Ltd

Griffiths, R. (1984). Consistent histories and the interpretation of quantum mechanics. *J. Stat. Phys.*, **36**, 219

Grossberg, S. (ed.)(1987). *The adaptive brain I: Cognition, learning, reinforcement and rhythm* and *The adaptive brain II: Vision, speech, language and motor control.* North-Holland, Amsterdam

Grünbaum, B. and Shephard, G.C. (1987). *Tilings and Patterns.* Freeman, New York

Grundler, W. and Keilmann, F. (1983). Sharp resonances in yeast growth proved nonthermal sensitivity to microwaves. *Phys. Rev. Letts.*, **51**, 1214-1216

Guccione, S. (1993). Mind the truth: Penrose's new step in the Gödelian argument. *Behavioural and Brain Sciences*, **16**, 612-613

Haag, R. (1992). *Local quantum physics: fields, particles, algebras.* Springer-Verlag, Berlin

Hadamard, J. (1945). *The psychology of invention in the mathematical field.* Princeton University Press

Hallett, M. (1984). *Cantorian set theory and limitation of size.* Clarendon Press, Oxford

Hameroff, S.R. (1974). Chi: a neural hologram? *Am. J. Clin. Med.*, **2(2)**, 163-170

Hameroff, S.R. (1987). *Ultimate computing. Biomolecular conciousness and nanotechnology,* North-Holland, Amsterdam

Hameroff, S.R. and Watt, R.C. (1982). Information in processing in microtubles. *J. Theor. Biol.*, **98**, 549-561

Hameroff, S.R. and Watt, R.C. (1983). Do anesthetics act by altering electron mobility? *Anesth. Analg.*, **62**, 936-940

Hameroff, S.R., Rasmussen, S. şi Mansson, B. (1988). Molecular automata in microtubles: basic computational logic of the living state? In *Artificial Life, SFI studies in the sciences of complexity* (ed. C. Langton), Addison-Wesley, New-York

Hanbury Brown, R. and Twiss, R.Q. (1954). A new type of interferometer for use in radio astronomy. *Phil. Mag.*, **45**, 663-682

Hanbury Brown, R. and Twiss, R.Q. (1956). The question of correlation between photons in coherent beams of light. *Nature*, **177**, 27-29

Harel, D. (1987). *Algorithmics. The spirit of computing.* Addison-Wesley, New York

Hawking, S.W. (1975). Particle creation by Black Holes. *Commun. Math. Phys.*, 43, 199-220

Hawking, S.W. (1982). Unpredictability of quantum gravity. *Commun. Math. Phys.*, 87, 395-415

Hawking, S.W. and Israel, W. (ed.) (1987). *300 years of gravitation.* Cambridge University Press

Hebb, D.O. (1949). *The organization of behaviour.* Wiley, New York

Hecht, S., Schlaer, S. şi Pirenne, M.H. (1941). Energy, quanta and vision. *Journal of General Physiology,* **25**, 891-940

Herbert, N. (1993). *Elemental mind. Human consciousness and the new physics.* Dutton Books, Penguin Publishing

Heyting, A. (1956). *Intuitionism: an introduction.* North-Holland, Amsterdam

Heywood, P. and Redhead, M.L.G. (1983). Nonlocality and the Kochen-Specker Paradox. *Found. Phys.,* **13**, 481-499

Hodges, A.P. (1983). *Alan Turing: the enigma.* Burnett Books and Hutchinson, London; Simon and Schuster, New York

Hodgkin, D. and Houston, A.I. (1990). Selecting for the con in consciousness. *Behaviuoral and Brain Sciences,* **13(4)**, 668

Hodgson, D. (1991). *Mind matters: consciousness and choice in a quantum world.* Clarendon Press, Oxford

Hofstadter, D.R. (1979). *Gödel, Escher, Bach: an eternal golden braid.* Harvester Press, Hassocks, Essex

Hofstadter, D.R. (1981). A conversation with Einstein's brain. In *The mind's I* (ed. D.R. Hofstadter and D. Dennett), Basic Books; Penguin, Harmondsworth, Middlesex

Hofstadter, D.R. and Dennett, D. (ed.) (1981) *The mind's I.* Basic Books; Penguin, Harmondsworth, Middlesex

Home, D. (1994). A proposed new test of collapse-induced quantum nonlocality. Preprint

Home, D. and Nair, R. (1994). Wave function collapse as a nonlocal quantum effect. *Phys. Lett.,* **A187**, 224-226

Home, D. and Selleri, F. (1991). Bell's Theorem and the EPR Paradox. *Rivista del Nuovo Cimento*, **14**, nr.9

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.,* **79**, 2554-2558

Hsu, F.-H., Anantharaman, T., Campbell, M. şi Nowatzyk, A. (1990). A grandmaster chess machine. *Scientific American*, **263**

Huggett, S.A. and Tod, K.P. (1985). *An introduction to twistor theory.* London Math. Soc. student texts, Cambridge University Press

Hughston, L.P., Jozsa, R. şi Wootters, W.K. (1993). A complete classification of quantum ensembles having a given density matrix. *Phys. Letters*, **A183**, 14-18

Isham, C.J. (1989). Quantum gravity. In *The new physics* (ed. P.C.W. Davies), p.70-93, Cambridge University Press

Isham, C.J. (1994). Prima facie questions in quantum gravity. In *Canonical relativity: classical and quantum* (ed. J. Ehlers and H. Friedrich), Springer-Verlag, Berlin

Jibu, M., Hagan, S., Pribram, K., Hameroff, S.R.,and Yasue, K. (1994). Quantum optical

coherence in cytoskeletal microtubles: implications for brain function. *Bio. Systems* (in press)

Johnson-Laird, P.N. (1983). *Mental models.* Cambridge University Press

Johnson-Laird, P.N. (1987). How could consciousness arise from the computations of the brain? In *Mindwaves: thoughts on intelligence, identity and consciousness* (ed. C. Blakemore and S. Greenfield), Blackwell, Oxford

Károlyházy, F. (1966). Gravitation and quantum mechanics of macroscopic bodies. *Nuo. Cim. A*, **42**, 390-402

Károlyházy, F. (1974). Gravitation and quantum mechanics of macroscopic bodies. *Magyar Fizikai Polyoirat*, **12**, 24

Károlyházy, F., Frenkel, A. şi Lukács, B. (1986). On possible role of gravity on the reduction of the wave function. In *Quantum concepts in space and time* (ed. R. Penrose and C.J. Isham), Oxford University Press

Kasumov, A.Y., Kislov, N.A. şi Khodos, I.I. (1993). Can the observed vibration of a cantilever of supersmall mass be explained by quantum theory? *Microsc. Microanal. Microstruct.*, **4**, 401-406

Kentridge, R.W. (1990). Parallelism and pattern of thought. *Behavioural and Brain Sciences*, 13(4), 670

Khalfa, J. (ed.) (1994). *What is intelligence? The Darwin College lectures.* Cambridge University Press

Klarner, D.A. (1981). My life among the Polyominoes. In *The mathematical gardner* (ed. D.A. Klarner), Prindle, Weber and Schmidt, Boston MA and Wadsworth Int., Belmont CA

Kleene, S.C. (1952) *Introduction to metamathematics*. North-Holland, Amsterdam and van Nostrand, New York

Klein, M.V. and Furtak, T.E. (1986). *Optics.* 2nd edn. Wiley, New York

Kochen, S. and Specker, E.P. (1967). The problem of hidden variables in quantum mechanics. *J. Math. Mech.*, **17**, 59-88

Kohonen, T. (1984). *Self-organization and associative memory.* Springer-Verlag, New York

Komar, A.B. (1969). Qualitative features of quantized gravitation. *Int. J. Theor. Phys.*, **2**, 157-160

Koruga, D. (1974). Microtuble screw symmetry: packing of spheres as a latent bioinformation code. *Ann. NY Acad. Sci.*, **466**, 953-955

Koruga, D., Hameroff, S., Withers, J., Loutfy, R. şi Sundareshan, M. (1993). *Fullerese C_{60}. History, physics, nanobiology, nanotechnology.* North-Holland, Amsterdam

Kosko, B. (1994). *Fuzzy thinking: the new science of fuzzy logic.* Harper Collins, London

Kreisel, G. (1960). Ordinal logics and the characterization of informal concepts of proof. *Proc. of the Internat. Cong. of Mathematics, Aug. 1958*, Cambridge University Press

Kreisel, G. (1967). Informal rigour and completeness proofs. In *Problems in the philosophy of mathematics* (ed. I. Lakatos), p. 138-186, North-Holland, Amsterdam

Laguës, M., Xiao Ming Xie, Tebbji, H., Xiang Zhen Xu, Mairet, V., Hatterer, C., et al. (1993). Evidence suggesting superconductivity at 250K in a sequentially deposited cuprate film. *Science*, **262**, 1850-1851

Lander, I.J. and Parkin, T.R. (1966). Couterexample to Euler's conjecture on sums of like powers. *Bull. Amer. Math. Soc.*, **72**, 1079

Leggett, A.J. (1984). Schrödinger's cat and her laboratory cousins. *Contemp. Phys.*, **25(6)**, 583

Lewis, D. (1969). Lucas against mechanism. *Philosophy*, **44**, 231-233

Lewis, D. (1989). Lucas against mechanism II. *Can. J. Philos.*, **9**, 373-376

Libet, B. (1990). Cerebral processes that distinguish conscious experience from unconscious mental functions. In *The principles of design and operation of the brain* (ed. J.C. Eccles and O.D. Creutzfeldt), Experimental Brain research series 21, p. 185-205, Springer Verlag, Berlin

Libet, B. (1992). The neural time-factor in perception, volition and free will. *Revue de Métaphysique et de Morale*, **2**, 255-272

Libet, B., Wright, E.W., Jr., Feinstein, B. şi Pearl, D.K. (1979). Subjective referral of the timing for a concious sensory experience. *Brain*, **102**, 193-224

Linden, E. (1993). Can animals think? *Time Magazine* (March) 13.

Lisboa, P.G.J. (ed.) (1992). *Neural networks: current applications*. Chapman Hall, London

Lockwood, M. (1989). *Mind, brain and the quantum*. Blackwell, Oxford

Longair, M.S. (1993). Modern cosmology - a critical assessment. *Q. J. R. Astr. Soc.*, **34**, 157-199

Longuet-Higgins, H.C. (1987). *Mental processes: studies in cognitive science*, Part II, MIT Press, Cambridge, Massachusetts

Lucas, J.R. (1961). Minds, machines and Gödel. *Philosophy*, **36**, 120-124; retipărit în Alan Ross Anderson (ed.) (1964) *Minds and Machines*, Englewood Cliffs

Lucas, J.R. (1970). *The freedom of the will*. Oxford University Press

McCarthy, J. (1979). Ascribing mental quatities to machines. In *Philosophical perspectives in artificial intelligence* (ed. M. Ringle), Humanities Press, New York

McCulloch, W.S. and Pitts, W.H. (1943). A logical calculus of the idea immanent in nervous activity. *Bull. Math. Biophys.*, **5**, 115-133 (Retipărit în McCulloch, W.S., *Embodiments of mind*, MIT Press, 1965)

McDermott, D. (1990). Computation and consciousness. *Behavioural and Brain Sciences,* **13(4)**, 676

MacLennan, B. (1990). The discomforts of dualism. *Behavioural and Brain Sciences,* **13(4)**, 673

Majorana, E. (1932). Atomi orientati in campo magnetico variable. *Nuovo Cimento,* **9**, 43-50

Manaster-Ramer, A., Savitch, W.J. şi Zadrozny, W. (1990). Gödel redux. *Behavioural and Brain Sciences,* **13(4)**, 675

Margulis, L. (1975). *Origins of eukaryotic cells.* Yale University Press, New Haven, CT

Markov, A.A. (1958). The insolubility of the problem of homeomorphy. *Dokl. Akad. Nauk. SSSR,* **121**, 218-220

Marr, D.E. (1982). *Vision: a computational investigation into the human representation and processing of visual information.* Freeman, San Francisco

Marshall, I.N. (1989). Consciousness and Bose-Einstein condensates. *New Ideas in Psychology,* **7**

Mermin, D. (1985). Is the moon there when nobody looks? Reality and the quantum theory. *Physics Today,* **38**, 38-47

Mermin, D. (1990). Simple unified form of the major no-hidden-variables theorems. *Phys. Rev. Lett.,* **65**, 3373-3376

Michie, D. and Johnston, R. (1984). *The creative computer. Machine intelligence and human knowledge.* Viking Penguin

Minsky, M. (1968). Matter, mind and models. In *Semantic information processing* (ed. M. Minsky), MIT Press, Cambridge, Massachusetts

Minsky, M. (1986). *The society of mind.* Simon and Schuster, New York

Minsky, M. and Papert, S. (1972). *Perceptrons: an introduction to computational geometry.* MIT Press, Cambridge, Massachusetts

Misner, C.W., Thorne, K.S. şi Wheeler, J.A. (1973). *Gravitation.* Freeman, New York

Moore, A.W. (1990). *The infinite.* Routledge, London

Moravec, H. (1988). *Mind children: the future of robot and human intelligence.* Harvard University Press, Cambridge, Massachusetts

Moravec, H. (1994). *The Age of Mind: transcending the human condition through robots.* In press.

Mortensen, C. (1990). The powers of machines and minds. *Behavioural and Brain Sciences,* **13(4)**, 678

Mostowski, A. (1957). *Sentences undecidable in formalized arithmetic: an exposition of the theory of Kurt Gödel.* North-Holland, Amsterdam

Nagel, E. and Newman, J.R. (1958). *Gödel's proof.* Routledge and Kegan Paul

Newell, A. and Simon, H.A. (1976). Computer science as empirical enquiry: symbols and search. *Communications of the ACM*, **19**, 113-126

Newell, A., Young, R. şi Polk, T. (1993). The approach through symbols. in *The simulation of human intelligence* (ed. D. Broadbent), Blackwell, Oxford

Newton, I. (1687). *Philosophiae Naturalis Principia Mathematica.* Reprint: Cambridge University Press

Newton, I. (1730). *Opticks.* 1952, Dover, New York

Oakley, D.A. (ed.) (1985). *Brain and mind.* Methuen, London

Obermayer, K., Teich, W.G. şi Mahler, G. (1988a). Structural basis of multistationary quantum systems. I. Effective single-particle dynamics. *Phys. Rev.*, **B37**, 8096-8110

Obermayer, K., Teich, W.G. şi Mahler, G. (1988b). Structural basis of multistationary quantum systems. II. Effective few-particle dynamics. *Phys. Rev.*, **B37**, 8111-8121

Omnès, R. (1992). Consistent interpretations of quantum mechanics. *Rev. Mod. Phys.*, **64**, 339-382

Pais, A. (1991). *Neils Bohr's times.* Clarendon Press, Oxford

Pauling, L. (1964). The hydrate microcrystal theory of general anesthesia. *Anesth.Analg.*, **43**, 1

Paz, J.P. and Zurek, W.H. (1993). Environment induced-decoherence, classicality and consistency of quantum histories. *Phys. Rev.*, **D48(6)**, 2728-2738

Paz, J.P., Habib, S. and Zurek, W.H. (1993). Reduction of the wave packet: preferred observable and decoherence time scale. *Phys. Rev.*, **D47(2)**, 3rd Series, 488-501

Pearle, P. (1976). Reduction of the state-vector by a nonlinear Schrödinger equation. *Phys. Rev.*, **D13**, 857-868

Pearle, P. (1989). Combining stochastic dynamical state-vector reduction with spontaneous localization. *Phys. Rev.*, **A39**, 2277-2289

Pearle, P. (1992). Relativistic model statevector reduction. In *Quantum chaos-quantum measurement*, NATO Adv.Sci.Inst.Ser.C. Math.Phys.Sci.358 (Copenhagen 1991), Kluwer, Dordrecht

Peat, F.D. (1988). *Superstrings and the search for the theory of everything.* Contemporary Books, Chicago

Penrose, O. (1970). *Foundations of statistical mechanics: a deductive treatment.* Pergamon, Oxford

Penrose, O. and Onsager, L. (1956). Bose-Einstein condensation and liquid helium. *Phys. Rev.*, **104**, 576-584

Penrose, R. (1980). On Schwarzschild causality - a problem for 'Lorenz covariant' general relativity. In *Essays in general relativity* (A. Taub Festschrift)(ed. F.J. Tipler),

p. 1-12, Academic Press, New York

Penrose, R. (1987). Newton, quantum theory and reality. In *300 Years of gravity* (ed. S.W. Hawking and W. Israel), Cambridge University Press

Penrose, R. (1990). Author's response, *Behavioural and Brain Sciences,* **13**(4), 692

Penrose, R. (1991a). The mass of the classical vacuum. In *The philosophy of vacuum* (ed. S. Saunders and H. R. Brown), Clarendon Press, Oxford

Penrose, R. (1991b). Response to Tony Dodd's 'Gödel, Penrose, and the possibility of AI' *Artificial Intelligence Review,* **5**, 235

Penrose, R. (1993a). Gravity and quantum mechanics. In *General relativity and gravitation 1992. Proceedings of the Thirteenth International Conference on General Relativity and Gravitation held at Cordoba, Argentina 28 June 4 July 1992. Part 1: Plenary lectures.* (ed. R.J. Gleiser, C.N. Kozameh, and O.M. Moreschi), Institute of Physics Publications, Bristol

Penrose, R. (1993b). Quantum non-locality and complex reality. On *The Renaissance of general relativity* (in honour of D.W. Sciama) (ed. G. Ellis, A. Lanza, and J. Miller), Cambridge University Press

Penrose, R. (1993c). Setting the scene: the claim and the issues. In *The simulation of human intelligence* (ed. D. Broadbent), Blackwell, Oxford

Penrose, R. (1993d). An emperor still without mind, *Behavioural and Brain Sciences,* **16**, 616-622

Penrose, R. (1994a). On Bell non-locality without probabilities: some curious geometry. In *Quantum reflections* (in honour of J.S.Bell) (ed. J. Ellis and A. Amati), Cambridge University Press

Penrose, R. (1994b). Non-locality in and objectivity in quantum state reduction. In *Fundamental aspects of quantum theory* (ed. J. Anandan and J.L. Safko), World Scientific, Singapore

Penrose, R. and Rindler, W. (1986). *Spinors and space-time,* Vol.2: *Spinor and twistor methods in space-time geometry.* Cambridge University Press

Percival, I.C. (1994). Primary state diffusion. *Proc. R. Soc. Lond.,* **A**, submitted.

Peres, A. (1985). Reversible logic and quantum computers. *Phys. Rev.,* **A32**(6), 3266-3276

Peres, A. (1990). Incompatible results of quantum measurements. *Phys. Lett.,* **A151**, 107-108

Peres, A. (1991). Two simple proofs of the Kohen-Specker theorem. *J. Phys. A: Math. Gen.,* **24**, L175-L178

Perlis, D. (1990). The emperor's old hat. *Behavioural and Brain Sciences,* 13(4), 680

Planck, M. (1906). *The thory of heat radiation* (trans. M. Masius, based on lectures delivered in Berlin, 1906/1907), 1959, Dover, New York

Popper, K.R. and Eccles, J.R. (1977). *The self and its brain.* Springer International

Post, E.L. (1936). Finite combinatory processes-formulation I, *Jour. Symbolic Logic*, **1**, 103-105

Poundstone, W. (1985). *The recursive universe: cosmic complexity and the limits of scientific knowledge.* Oxford University Press

Pour-El, M.B. (1974). Abstract computability and its relation to the general purpose analog computer. (Some connections between logic, defferential equations and analog computers) *Trans. Amer. Math. Soc.*, **119**, 1-28

Pour-El, M.B. and Richards, I. (1979). A computable ordinary differential equation which possesses no computable solution. *Ann Math.Logic*, **17**, 61-90

Pour-El, M.B. and Richards, I. (1981). The wave equation with computable initial data such that its unique solution is not computable. *Adv. in Math.*, **39**, 215-239

Pour-El, M.B. and Richards, I. (1982). Noncomputability in models of physical phenomena. *Int. J. Theor. Phys.*, **21**, 553-555

Pour-El, M.B. and Richards, I. (1989). *Computability in analysis and physics.* Springer-Verlag, Berlin

Pribram, K.H. (1966). Some dimensions of remembering: steps toward a neuropsychological model of memory. In *Macromolecules and behaviour* (ed. J. Gaito), p. 165-187, Academic Press, New York

Pribram, K.H. (1975). Toward a holonomic theory of perception. In *Gestalttheorie in der modern psychologie* (ed. S. Ertel), p. 161-184, Erich Wengenroth, Köln

Pribram, K.H. (1991). *Brain and perception: holonomy and structure in figural processing,* Lawrence Erlbaum Assoc., New Jersey

Putnam, H. (1960). Minds and machines. In *Dimensions of mind* (ed. S. Hook), New York symposium. Retipărit în *Minds and machines* (ed. A.R. Anderson), p. 43-59, Prentice-Hall, 1964; şi deasemeni în *Dimensions of mind: a symposium (Proceedings of the third annual NYU Institute of Philosophy),* p. 148-179, NYU Press, 1964

Ramon y Cajal, S. (1955). *Studies on the cerebral cortx* (trans. L.M. Kroft), Lloyd-Luke, London

Redhead, M.L.G. (1987). *Incompleteness, nonlocality, and realism.* Clarendon Press, Oxford

Rosenblatt, F. (1962). *Principles of neurodynamics.* Spartan Books, New York

Roskies, A. (1990). Seeing truth or just seeming true? *Behavioural and Brain Sciences,* **13**(4), 682

Rosser, J.B. (1936). Extensions of some theorems of Gödel and Church. *Jour. Symbolic Logic,* **1**, 87-91

Rubel, L.A. (1985). The brain as an analog computer. *J. Theoret. Neurobiol.*, 4, 73-81

Rubel, L.A. (1988). Some mathematical limitations of the general-purpose analog computer. *Adv. in Appl. Math.*, **9**, 22-34

Rubel, L.A. (1989). Digital simulation of analog computation and Church's thesis. *Jour. Symb. Logic*, **54(3)**, 1011-1017

Rucker, R. (1984). *Infinity and the mind: the science and philosophy of the infinite*. Paladin Books, Granada Publishing Ltd., London (First published by Harvester Press Ltd., 1982)

Sacks, O. (1973). *Awakenings*. Duckworth, London

Sacks, O. (1985). *The man who mistook his wife for a hat*. Duckworth, London

Sagan, L. (1967). On the origin of mitosing cells. *J. Theor. Biol.*, **14**, 225-274

Sakharov, A.D. (1967). Vacuum quantum fluctuations in curved space and the theory of gravitation. *Doklady Akad. Nauk SSSR*, 177, 70-71. Traducere în engleză: *Sov. Phys. Doklady*, **12**, 1040-1041 (1968)

Schrödinger, E. (1935a). 'Die gegenwartige Situation in der Quantenmechanik', *Naturwissenschaftenp*, **23**, 807-812, 823-828, 844-849 (Traducere în engleză de J.T. Timmer (1980) în *Proc. Amer. Phil. Soc.*, **124**, 323-338). In *Quantum theory and measurement* (ed. J.A. Wheeler and W.H. Zurek), Princeton University Press, 1983

Schrödinger, E. (1935b). Probability relations between separated systems. *Proc. Camb. Phil. Soc.*, **31**, 555-563

Schrödinger, E. (1967). *'What is life?' and 'Mind and Matter'* Cambridge University Press

Schroeder, M. (1991). *Fractals, chaos, power laws. Minutes from an infinite paradise*. Freeman, New York

Scott, A.C. (1973). Information processing in dendritic trees. *Math. Bio. Sci.*, **18**, 153-160

Scott, A.C. (1977). *Neurophysics*. Wiley Interscience, New York

Searle, J.R. (1980). Minds, brains and programs. In *The behavioral and brain sciences*, Vol.3, Cambridge University Press (Retipărit în *The mind's I* (ed. D.R. Hofstadter and D.C. Dennett), Basic Books, Inc.; Penguin Books Ltd., Harmondworth, Middlesex, 1981)

Searle, J.R. (1992). *The rediscovery of the mind*. MIT Press, Cambridge, Massachusetts

Seymore, J. and Norwood, D. (1993). A game for life. *New Scientists*, 139, nr.1889, 23-26

Sheng, D., Yang, J., Gong, C. şi Holz, A. (1988). A new mechanism of high Tc superconductivity. *Phys. Lett.*, **A133**, 193-196

Sloman, A. (1992). The emperor's real mind: review of Roger Penrose's The Emperor's new Mind. *Artificial Intelligence*, 56, 355-396

Smart, J.J.C. (1961). Gödel's theorem, Church's theorem and mechanism. *Synthèse*, **13**, 105-110

Smith, R.J.O. and Stephenson, J. (1975). *Computer simulation of continuous systems*. Cambridge University Press

Smith, S., Watt, R.C. şi Hameroff, S.R. (1984). Cellular automata in cytoskeletal lattice proteins. *Physica D*, **10**, 168-174

Smolin, L. (1993). What have we learned from non-perturbative quantum gravity? In *General relativity and gravitation 1992. Proceedings of the thirteenth international conference on GRG, Cordoba, Argentina* (ed. R.J. Gleiser, C.N. Kozameh, and O.M.Moreschi), Institute of Physics Publications, Bristol

Smolin, L. (1994). Time, structure and evolution in cosmology. In *Temponelle scieziae filozofia* (ed. E. Agazzi), World Scientific, Singapore

Smorynski, C. (1975). *Handbook of mathematical logic*. North-Holland, Amsterdam

Smorynski, C. (1983). 'Big' news from Archimedes to Friedman. *Notices Amer. Math. Soc.*, **30**, 251-256

Smullyan, R. (1961). *Theory of Formal Systems*. Princeton University Press

Smullyan, R. (1992). *Gödel's incompleteness theorem*, Oxford Logic Guide No.19, Oxford University Press

Squires, E.J. (1986). *The mystery of the quantum world*. Adam Hilger Ltd., Bristol

Squires, E.J. (1990). On an alleged proof of the quantum probability law. *Phys. Lett.*, **A145**, 67-68

Squires, E.J. (1992a). Explicit collapse and superluminal signals. *Phys. Lett.*, **A163**, 356-358

Squires, E.J. (1992b). History and many-worlds quantum theory. *Found. Phys. Lett.*, **5**, 279-290

Stairs, A. (1983). Quantum logic, realism and value-definiteness. *Phil. Sci.*, **50(4)**, 578-602

Stapp, H.P. (1979). Whiteheadian approach to quantum theory and the generalized Bell's theorem. *Found. Phys.*, **9**, 1-25

Stapp, H.P. (1993). *Mind, matter, and quantum mechanics*. Springer-Verlag, Berlin

Steen, L.A. (ed.) (1978). *Mathematics today: twelve informal essays*. Springer-Verlag, Berlin

Stoney, G.J. (1881). On the physical units of nature. *Phil. Mag.*, (Series 5)**11**, 381

Stratton, A.O.W, Davis, R.E., Angstadt, J.D., Donmoyer, J.E., Johnson, C.D. şi Meade, J.A. (1987). Nematode neurobiology using Ascaris as a model system. *J. Cellular Biochem.*, **511A**, 144

Thorne, K.S. (1994). *Black holes & time warps: Einstein's outrageous legacy*. W.W Norton and Company, New York

Torrence, J. (1992). *The concept of nature. The Herbert Spencer lectures,* Clarendon Press, Oxford

Tsotsos, J.K. (1990). Exactly which emperor is Penrose talking about? *Behavioural and Brain Sciences,* **13(4)**, 686

Turing, A.M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proc. London. Math. Soc., (ser.2),* **42**, 230-265; a correction **43**, 544-546

Turing, A.M. (1939). Systems of logic based on ordinals. *P. London. Math. Soc.,* **45**, 161-228

Turing, A.M. (1950). Computing machinery and intelligence. *Mind,* **59** No.236. Retipărit în *The min's I* (ed. D.R. Hofstadter and D.C. Dennett), Basic Books; Penguin, Harmondsworth, Middlesex, 1981

Turing, A.M. (1986). Lecture to the London Mathematical Society on 20 February 1947. In *A.M. Turing's ACE report of 1946 and other papers* (ed. B.E. Carpenter and R.W. Doran), The Charles Babbage Institute, vol. 10, MIT Press, Cambridge, Massacusetts

von Neumann, J. (1932). *Mathematische Grundlagen der Quantummechanik,* Springer-Verlag, Berlin; traducere în engleză: *Mathematical foundations of quantum mechanics,* Princeton University Press, 1955

von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economoc behaviour.* Princeton University Press, 1955

Waltz, D.L. (1982). Artificial Intelligence. *Scientific American,* **247**(4), 101-122

Wang, Hao (1974). *From mathematics to philosophy,* Routledge, London

Wang, Hao (1987). *Reflections on Kurt Gödel.* MIT Press, Cambridge, Massachusetts

Wang, Hao (1993). On physicalism and algorithmism; can machines think? *Philosophia mathematica* (Ser. III), p. 97-138

Ward, R.S. and Wells, R.O.Jr. (1990). *Twistor geometry and field theory,* Cambridge University Press

Weber, J. (1960). Detection and generation of gravitational waves. *Phys. Rev.,* **117**, 306

Weinburg, S. (1977). *The first three minutes: a modern view of the origin of the universe.* Andre Deutsch, London

Wheeler, J.A. (1957). Assesment of Everett's relative state' formulation of quantum theory. *Revs. Mod. Phys.,* **29**, 463-465

Wheeler, J.A. (1975). On the nature of quantum geometrodynamics. *Annals of Phys.,* **2**, 604-614

Wigner, E.P. (1960). The unreasonable effectiveness of mathematics. *Common. Pure Appl. Math.,* **13**, 1-14

Wigner, E.P. (1961). Remarks on the mind-body equation. In *The scientist speculates* (ed. I.J. Good), Heinemann, London (Retipărit în E. Wigner (1967), *Symmetries and reflections,* Indiana University Press, Bloomington; şi în *Quantum theory and measurement* (ed. J.A. Wheeler and W.H. Zurek) Princeton University Press, 1983)

Wilensky, R. (1990). Computability, consciousness and algorithms. *Behavioural and Brain Sciences,* **13(4)**, 690

Will, C. (1988). *Was Einstein right? Putting general relativity to the test.* Oxford University Press

Wolpert, L. (1992). *The unnatural nature of science.* Faber and Faber, London

Woolley, B. (1992). *Virtual worlds.* Blackwell, Oxford

Wykes, A. (1969). *Doctor Cardano. Physician extraordinary.* Frederich Muller

Young, A.M. (1990). *Mathematics, physics and reality.* Robert Briggs Associates, Portland, Oregon

Zeilinger, A., Gaehler, R., Shull, C.G. şi Mampe, W. (1988). Single and double slit diffraction of neutrons. Revs. Mod. Phys., **60**, 1067

Zeilinger, A., Horne, M.A. şi Greenberger, D.M. (1992). Higher-order quantum entanglement. In *Squeezed states and quantum uncertainty* (ed. D. Han, Y.S. Kim, and W.W.Zachary), NASA Conf. Publ. 3135, NASA, Washington, DC

Zeilinger, A., Zukowski, M., Horne, M.A., Bernstein, H.J. şi Greenberger, D.M. (1994). Eistein-Podolsky-Rosen correlations in higher dimensions. In *Fundamental aspects of quantum theory* (ed. J. Anandan and J.L. Safko), World Scientific, Singapore

Zimba, J. (1993). Finitary proofs of contextuality and nonlocality using Majorana representation of spin-3/2 states, M.Sc. thesis, Oxford

Zimba, J. and Penrose, R. (1993). On Bell non-locality without probabilities: more curious geometry. *Stud. Hist. Phil. Sci.,* **24(5)**, 697-720

Zohar, D. (1990). *The quantum self. Human nature and consciousness defined by the New Physics.* William Morrow and Company, Inc., New York

Zohar, D. and Marshall, I. (1994) *The quantum society. Mind, physics and a new social vision.* Bloomsbury, London

Zurek, W.H. (1991). Decoherence and the transition from quantum to classical. *Physics Today,* **44** (Nr.10), 36-44

Zurek, W.H. (1993). Preferred states, predictability, classicality and the environment-induced decoherence. *Prog. of Theo. Phys.,* **89(2)**, 281-302

Zurek, W.H., Habib, S. şi Paz, J.P. (1993). Coherent states via decoherence. *Phys. Rev. Lett.,* **70(9)**, 1187-1190

Sir Roger Penrose : scurtă notă biografică

Roger Penrose este profesor de matematică, titularul catedrei „Rouse Ball" de la Universitatea din Oxford.

S-a născut în 1931, în Colchester, Essex, Marea Britanie. Licenţiat în matematică la „University College" din Londra. Doctoratul în geometrie algebrică l-a obţinut la Cambridge, la „St. John's College". A ocupat mai multe posturi în Marea Britanie şi Statele Unite, în particular la „Birkbeck College", Londra.

Provine dintr-o familie cu tradiţie ştiinţifică şi afinităţi matematice, cu părinţi medici, tatăl genetician cu preocupări matematice deosebite, cu un frate matematician şi un altul - 10 ani campion naţional la şah. În zona artelor şi în psihologia vederii, numele său este asociat *scării lui Penrose* şi triunghiului imposibil numit *tribar* (create de Roger împreună cu tatăl său), figuri imposibile, folosite în celebrele desene ale lui M.C. Escher.

Contribuţiile sale în cercetare includ geometria - cu teoria acoperirilor neperiodice sau de diviziune a planului (non-periodic *tilings)*, teoria relativităţii şi fundamentele teoriei cuantice, dar şi ştiinţa conştiinţei (*science of consciousness*). Programul său principal de cercetare vizează elaborarea unei teorii a *twistor*-ilor, lansată de el cu peste 30 de ani în urmă ca o tentativă de unificare a teoriei relativităţii generalizate (a lui Einstein) cu mecanica cuantică.

A obţinut numeroase premii şi titluri, dintre care trebuie menţionate:
- Premiul Wolf pe 1988, împreună cu fizicianul Stephen Hawking pentru studiile asupra Universului,
- Medalia „Royal" a Societăţii Regale,
- Premiul Albert Einstein.

Cartea sa din 1989 „The Emperor's New Mind", devenită best-seller, a fost premiată cu premiul cărţii de ştiinţă pe 1990 „'Rhone-Poulenc' Science Book Prize" Cartea de faţă, „Shadows of the Mind" a apărut în 1994 şi a stârnit controverse şi mai serioase decât precedenta, controverse ce au condus şi la apariţia în 1996 a unei societăţi în Internet numită „Sir Roger Penrose Society" (SRPS) (http://www.liu.se/srps/). Tot în 1994 a fost înobilat (cu titlul de *Sir*) pentru serviciile aduse ştiinţei.

Termeni propuşi în limba română

Au fost aleşi pentru unii dintre termenii englezeşti următorii echivalenţi în limba română:

assailable/ unassailable	atacabil/ de neatacat, de nezdruncinat
assert	a afirma
awareness	conştienţă
belief	convingere
computational/ non-computational	computaţional non-computaţional
consciousness	conştiinţă
consciousness awareness	conştiinţă conştientă
consistent	necontradictoriu, consistent
insight	intuiţie
knowable/ unknowable	cognoscibil, care poate fi înţeles/ incognoscibil, care nu poate fi înţeles
loophole	portiţă de scăpare
mind	1) mintea umană ca întreg (conştient + inconştient, subiect, uneori suport fizic) 2)intelectul uman (suma proceselor psihice superioare care au loc în creier)
mind body problem	(de la Descartes: legătura sau separaţia între *substanţa gânditoare* şi *substanţa extinsă*)
perceive	a percepe, a înţelege
quantum entaglement	non-localizare cuantică
random	aleator
scientific world view	imaginea ştiinţifică a lumii
sound	corect
sound formal system	sistem formal în care expresiile care pot fi derivate sintactic sunt adevărate
sound reasoning	raţionament corect
tilted	înclinat
underlie	a sta la baza

A mai apărut:

ROGER PENROSE

MINTEA NOASTRĂ... CEA DE TOATE ZILELE

despre gândire, fizică şi calculatoare

În *Mintea noastră cea de toate zilele*, o lucrare îndrăzneață, scânteietoare, ce încearcă să depăşească bazele existente, autorul demonstrează că nouă ne lipsesc în prezent elemente importante de înțelegere a fundementelor fizicii, fără de care nu vom putea înțelege vreodată gândirea..."

The New York Times
Book Review

În curând se va reedita:

FRITJOF CAPRA

TAOFIZICA

Lucrarea reprezintă o inedită explorare a paralelismelor dintre fizica modernă şi filosofiile orientale, avansând ideea conform căreia tradiţiile misticii orientale, pot constitui un cadru filosofic coerent celor mai avansate teorii din fizică.

În pregătire:

JOHN D. BARROW

DESPRE IMPOSIBILITATE
LIMITELE ŞTIINŢEI – ŞTIINŢA LIMITELOR

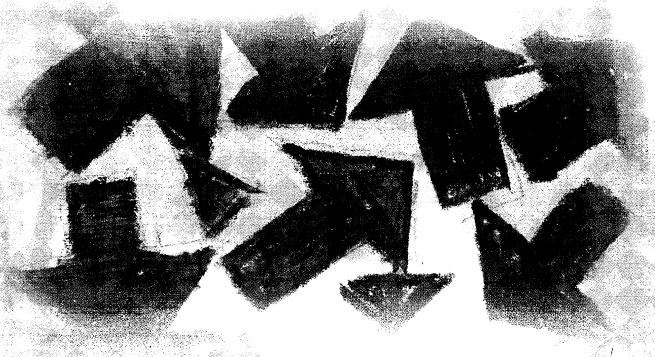
Cartea lui J. Barrow se ocupă de una dintre cele mai tulburătoare întrebări ale conştiinţei umane: există limite ale cunoaşterii lumii de către mintea umană? Autorul analizează cu multă pătrundere, într-un stil atractiv, cu exemple la îndemâna cititorului, aspectele multiple ale unei posibile limitări a ştiinţei.

O provocare pentru lumea ştiinţifică (care nu are o teorie despre conştiinţă), lucrarea se adresează, de fapt, unui public mult mai larg, tuturor celor interesaţi în întrebările fundamentale care ne marchează încă viaţa.

Penrose s-a apropiat mai mult decât oricine altcineva de o discuţie riguroasă a celei mai stringente dintre problemele noastre: *cine suntem?... cum gândim?... şi ce este acel ceva care ne face umani?...*

The Times

Tehnos



ISBN 973-31-1290-9